

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2005, 11:59:50 ; Search time 5661 Seconds
(without alignments)
12809.121 Million cell updates/sec

Title: US-10-757-093-3

Perfect score: 1905

Sequence: 1 atgaattcctcagggatt.....ctagatgataagaattaa 1905

Scoring table: IDENTITY NUC

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: gb_esc1:*
2: gb_esc2:*
3: gb_hic:*
4: gb_esc3:*
5: gb_esc4:*
6: gb_esc5:*
7: gb_esc6:*
8: gb_gsa1:*
9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	261.2	13.7	921	9	CL468845 SAIL_443
2	222	11.7	2274	3	AK041058 Mus muscu
3	208.6	11.0	566	7	CV169266 remexl_01
4	205.8	10.8	1995	6	CD014092 90134967
5	188.4	9.9	2473	3	AY321342 Rattus no
6	183.2	9.6	1377	3	CR593823 full-len
7	183.2	9.6	1853	6	CD014093 90135266
8	180.4	9.5	857	5	BK745933 BX745933
9	178.8	9.4	811	9	CL679241 PRI0125C
10	175.4	9.2	908	9	CL463714 SAIL_119
11	166.2	8.7	1124	6	CD503076 CDA60-C07
12	164.8	8.7	907	9	CL478429 SAIL_28 D
13	158.8	8.3	1138	6	CD503098 CDA60-D07
14	155.4	8.2	740	7	CF521612 AGENCOURT
15	150.6	7.9	689	7	CR442804 CR442804
16	145.6	7.6	914	6	CB203472 AGENCOURT
17	141.4	7.4	767	6	CF255373 agvnl27.c
18	139.2	7.3	906	5	BQ941196 AGENCOURT
19	139	7.3	957	5	CL495417 SAIL_60 A
20	137.8	7.2	1051	5	BK363460 BX363460
21	137.4	7.2	1268	6	CD503077 CDA60-C07
22	133.6	7.0	913	9	CL474264 SAIL_215
23	131	6.9	857	6	CB559443 AGENCOURT
24	130.6	6.9	952	9	CL468844 SAIL_443

25	129.4	6.8	2095	6	CD014094 90135027
26	127	6.7	907	9	CL482753 SAIL_367
27	126.6	6.7	624	4	BI445636 deeb3e05-
28	126.4	6.6	963	9	CL461373 SAIL_1146
29	126.2	6.6	892	9	CL462844 SAIL_1171
30	126	6.6	616	2	BE179118 R00-HT061
31	126	6.6	900	6	CA489317 AGENCOURT
32	125.8	6.6	790	7	CO559438 AGENCOURT
33	124.8	6.6	667	7	CN264045 170004243
34	123.8	6.5	884	9	CL461909 SAIL_1154
35	123.2	6.5	802	5	BUS37212 603474094
36	122.4	6.4	799	5	BUS36987 603585124
37	121.6	6.4	1023	5	BK363459 BX363459
38	121.2	6.4	816	7	CK467358 938649 MA
39	121	6.4	752	4	BJ733141 BJ733141
40	121	6.4	846	6	CA453907 AGENCOURT
41	120.6	6.3	791	5	BUS39978 603323761
42	120	6.3	603	1	AL676250 AL676250
43	119.4	6.3	725	5	BUS15788 603851444
44	119.4	6.3	777	4	BG741172 602631852
45	119	6.2	639	5	BM997809 UI-H-D10-

ALIGNMENTS

RESULT 1
LOCUS CL468845/C
DEFINITION SAIL_443_B06.v2 SAIL Collection Arabidopsis thaliana genomic clone
SAIT_443_B06.v2, genomic survey sequence.
ACCESSION CL468845
VERSION CL468845.1 GI:45965487
KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 921)

Authors: Sessions, A., Burke, E., Presting, G., Aux, G., McEliver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J.D., Cotton, D., Bullis, D., Snell, J., Miguel, T., Hutchinson, D., Kimmery, B., Mittel, T., Katagiri, F., Glazebrook, J., Law, M. and Goff, S.A.
Title: A high-throughput Arabidopsis reverse genetics system
Journal: Plant Cell 14 (12), 2985-2994 (2002)
Medline: 12468722
PubMed: 22356987

COMMENT Contact: Sessions A

Applied Trait Genetics
Syngenta Biotechnology Inc.
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
Email: allen.sessions@syngenta.com
ABRC Stock Number CS820387; T-DNA left border flanking sequences of
Syngenta Arabidopsis insertion library (SAIL) lines are available
through the Arabidopsis Biological Resource Center (ABRC).
Sequences represent a pool of amplified genomic regions and not
single contiguous sequences.
Class: TDNA tagged.

FEATURES
source Location/Qualifiers
1..921
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="SAIL_443_B06.v2"
/clone_lib="SAIL Collection"
/note="T-DNA left border sequences were isolated using a
modified TAIL-PCR strategy"

ORIGIN

Query Match 13.7%; Score 261.2; DB 9; Length 921;
Best Local Similarity 61.2%; Pred. No. 1.6e-70;

Matches	442;	Conservative	0;	Mismatches	273;	Indels	7;	Gaps	1;
QY	1160	GAATTCGTCGATCGATGAACACCTGCCGTGATCGAATTCGCTTGATGGGGTAT	1219						
Db	795	GCATCTGTGATGATGATGAACCTGCTCTGCTGCTTAACTCTCTTGAGCATGGTT	736						
QY	1220	CTGAGAGTGTGCC-----CCACAACATTACGCCAGATCGATTACATTAAC	1272						
Db	735	TCGAGGGGGCAACAGCCGAAAGAACTGACAGGAAAGAGCAGTCAACGGGGAAC	676						
QY	1273	CNAGAGCCCAAGCAGAGGATTCGTGAGCTATTGCCGGAACAAACCATGCCAGT	1332						
Db	675	CACCAAGCCGACTTACAGGGGATTAAAGACTATAGCGGTACAAACCCCAAGC	616						
QY	1333	GTTGTCATGTGCTCTATTGCAACGAGCCGCATCTCATAGATGAGTCCCGAATAC	1392						
Db	615	GTGGTATGTGGAGTATGTCACCAAGACCGGATACCGTCCGAGGTGACGGGAAT	556						
QY	1393	TTGAGCCCATGACCAATTGACTCGTCAACTTGATCAACTGCCCTATTACATTTG	1452						
Db	555	TTGGCCGCACTGGCGGAAGACGCGTAACTGACCCGACGCGTCGATCAGCTG	496						
QY	1453	AAGTGGGACGCGGACATTCAGCTGATCGATCTCTGATCTGTTGATGTCATGTC	1512						
Db	495	AATGTATATGTTCTGCAACGCTCACCGATACATCAGCATCTTTGATGTCGTGC	436						
QY	1513	ATAATCGTATTTGATGATGATGATTTCTCAACAGAGACCTTGAGAGAGCAGG	1572						
Db	435	CTGAAACCGTTTATTCGATGATGATGATGATGATGATGATGATGATGATGAT	376						
QY	1573	CTTGAAGAGAGTGCATGATGATGATGATGATGATGATGATGATGATGATGAT	1632						
Db	375	CTGGAAGAGAGATCTGCGCTGCGGAGAGAACTGATGATGATGATGATGATGAT	316						
QY	1633	TATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1692						
Db	315	TATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	256						
QY	1693	TTTCAAGTACAAATGCTAGATGATGATGATGATGATGATGATGATGATGAT	1752						
Db	255	TATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	196						
QY	1753	GCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1812						
Db	195	GCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	136						
QY	1813	GGTAAAGAGAGTGTGTTTACCCGTGACCGGAAAGCGGAGCTCATAGTTG	1872						
Db	135	GGTAAAGAGAGATCTTCACTCGGACCGGAAAGCGGAGCTCATAGTTA	76						
QY	1873	AG 1874							
Db	75	AG 74							

RESULT 2
AK041058
LOCUS
DEFINITION
AK041058 2274 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male aorta and vein cDNA, RIKEN full-length
enriched library, clone:AS30072005 product:beta-glucuronidase
structural, full insert sequence.
AK041058
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Mus musculus (house mouse)
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
99279253

REFERENCE	PUBMED	10349636
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.	
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)	
MEDLINE	20493374	
PUBMED	11042159	
REFERENCE		
AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitanai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, B., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.	
TITLE	RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer	
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)	
MEDLINE	20530913	
PUBMED	11076861	
REFERENCE		
AUTHORS	4	
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.	
JOURNAL	Functional annotation of a full-length mouse cDNA collection	
REFERENCE	Nature 409, 685-690 (2001)	
AUTHORS	5	
TITLE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	
JOURNAL	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
REFERENCE	Nature 420, 563-573 (2002)	
AUTHORS	6 (bases 1 to 2274)	
ADACHI, J., AIZAWA, K., AKIMURA, T., ARAKAWA, T., BONO, H., CARNINCI, P., FUKUDA, S., FURUKAWA, M., HANAGAKI, T., HARA, A., HASEIZUME, W., HAYASHIDA, K., HAYASHI, N., HIRAMOTO, K., HIRAKAWA, I., HITOZANE, T., HORI, F., IMOTANI, K., ISHII, Y., ITOH, M., KAGAWA, T., KASHIWA, T., KATO, H., KAWAI, J., KOJIMA, Y., KONDO, S., KONO, H., KOWA, M., KOYA, S., KURIHARA, C., MATSUYAMA, T., MIYAZAKI, A., MURATA, M., NAKAMURA, M., NISHI, K., NOMURA, K., NUMAZAKI, R., OHNO, M., OHSATO, N., OKAZAKI, Y., SAITO, R., SAITO, H., SAKAI, C., SAKAI, K., SAKAZUME, N., SANO, H., SASAKI, D., SHIBATA, K., SHINGAWA, A., SHIRAKI, T., SOGABE, Y., TAGAMI, M., TAGAWA, A., TAKAHASHI, F., TAKAKU-AKANISHI, S., TAKEEDA, Y., TANAKA, T., TOMARU, A., TOYA, T., YASUNISHI, A., YAMAMOTO, M., and Hayashizaki, Y.		
TITLE	Direct Submission	
JOURNAL	Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Teikoku-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	
COMMENT	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
FEATURES	Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://phantom.gsc.riken.jp/ Location/Qualifiers 1..2274 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM:DB:A530072005" /db_xref="taxon:10090" /clone="A530072005" /sex="male" /issue_type="aorta and vein" /clone_lib="RIKEN full-length enriched mouse cDNA library"	

```

misc_feature      /dev_stage="adult"
                  14..1778
                  /note="beta-glucuronidase structural [MGDI:95874,
                  GB|NM_010368, evidence: BLASTN, 99%, match=733]
polyA_signal      putative"
                  2260..2265
polyA_site        /note="putative"
                  2274
                  /note="putative"

ORIGIN
Query Match      11.7%; Score 223; DB 3; Length 2274;
Best Local Similarity 51.3%; Pred. No. 6e-58;
Matches 695; Conservative 0; Mismatches 590; Indels 69; Gaps 5;

QY 563 GGAACCGCATGCGCAAGAAATCCAGACTATCAACATGACTTTTCAATGCTGTG 622
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 409 GGTATCCCAAGGGTTATCTTGTCCAGAGACAAGCTTTGACTTTCATGCGGGAC 468
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 623 TCGCCGATCTATCTGGCTTTTATTTCTGTACCCAGACAATATCCAGATATTACTGTG 682
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 469 TGCATGATCTGTGTCTCTATATCCACCCCTTACCTTACATGATGATATCATCTGTG 528
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 683 TTACAGATGTGATGTGCAATGTGTGATTTAACTACAGAGGTGCAAGTGGCGAACAGA 742
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 529 TCACATAATGTGAGCAAGCATCGGGCTGTGCTACTGGAATTTCTGTGAGGGCAGTG 588
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 743 CGACGGGGCAATCCAGATCTCAGTATGACAGAGATGAGATATTGTTGCAAGCCCT 802
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 589 AACATTTCCAGCTAGAAAGTCAACTTTTGTGATGAGGTGGCAAAAGTCGGCCCATGGGA 648
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 803 CGGAGCTCAGGGTACTGTCAAAATTCCTCAGTCAAGCTATGGCAACTGGCGCCGAT 862
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 649 CAGGGAACCGGGTCAACTTCAGGTTTCCAGTCCAACTCTGTGTGCTTACCTGATGC 708
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 863 ATCTTACCACTCCAGTCAAC-----ATCGTGGTCTTACCGCGGATGTAG 910
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 709 ATGAGCATCCAGCTCATGTACTCTGTGAGGTGAAGGTGAACAACATGAGTCTGTA 768
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 911 TCGACACCTTCAATTTGGCTACGGGGTGTCTACTGTCAAGTGTCCGGGTCACAATTC 970
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 769 CTGACTACTCAACCCCTTCTATCGGGATTCGAACATGTGCTGTCAAAAGCAAACTTCC 828
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 971 TATATAATGAAGCCTTTCTACTTTACCGGTTTTGGCAAACTGAAGCACAGCATAC 1030
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 829 TCAATAACGGAAAGCCCTTCTATTTCCAAAGGGGTCAATAGACAGAGATTCAATATCC 888
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1031 GTGGCAAGACATGACCCAGACATACATGTTTCAAGATTTCCAACTCATGAATGATTTG 1090
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 889 GAGGGAAGAGCTTTCGACCTGGCGGTGTGTAAGATTTCAACCTGCTCCGTGGCTGG 948
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1091 GAGCAATTTCTTTTGGCACTTCACTATTCCTTACCGGGAAGGTCAATGATTTGGCAG 1150
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 949 GGGCAAAATTCCTTCTTACCAAGCACTATCCCTACTCAAGAGGGTACTTCAGCTCTGTG 1008
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1151 ATCGAAATGGAATGTGCGATGATGAAGAAACCTGCGGTGGTCTGAACATTGCTTGA 1210
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1009 ACCGATACGGAGATGTGTGATCATGATGAGTGTCCCGGTGGG-----CA 1053
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1211 TGGGCGTATCTGAGATGTGCCCCCAAAACATTTTACGCCAGATGCAATTAACATTA 1270
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1054 TTGTGTACTCTAGAGTTTGGGAAGAGTCACTTGGC-----1092
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1271 CCCAAGAGCCCAAGACAGCGGATTCGTGAGTCACTTGGCCGAGCAAAAACCATCCA 1330
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1093 -----ACCACTAGAGGTATGAGAGAGCTGTGCGCCGGCAAAAATCACCTCG 1143
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1331 GTGTGTCAATGTGTCTATTTGCAAGAGCCGATCTCATGAAGATGAGTCTGGGAAT 1390
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1144 CGGTGTGATGTGTGTGTGTGGCCAAATGAGCTTCTCTGTCTGAAACCGCGCATATT 1203
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1391 ACTTGCAGCACTGACCAATTTGACTGTCAACTTGTATCAACTCGCCTATTATCAATTTG 1450
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 1204 ACTTTAAGCGCTGATCAACCCACCAACCCCTGAGCTCAACCCGTCCGTGACCTTT- 1262
QY 1451 CTAACTCGGCACGGGCAATATACAGCTGATGATCTCTGATCTGTTGATGTGAGTT 1510
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1263 -----GTAGCAACCCCAATATGATGACAGCTTGGGGGCCCCGTACGTGATATCT 1317
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1511 GCATAAATCGTATTTTGGATGTATTTCAAAACAGAGACCTTGAAGAGCAGAGCGAG 1570
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1318 GTGTAAACAGTACTACTTTCTGTGTATCATGACTATGGGCAATTTGAGGTGATTCAGCCAC 1377
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1571 CTCTGAAAAGAGCTGACATGATGCAAGAGAAATTCACAGCCGATCTCATACCG 1630
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1378 AGCTGAATGCCAGTTTGAACCTGTATTAAGACGATCAGAGCCGATTAATCCAGAGG 1437
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1631 AATATGTCAGATACCTCTTGAGAGGCTTCAGCTTATCTCTGAGACGCTTGGAGCGAAG 1690
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1438 AGTATGAGACAGAGCAATCCAGGATCAAGAGAACCCGCTGCATGTTCAATGAGG 1497
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1691 AGTTCCAGTACAAATGTGACATGTACATCAATGATGTTGAT-----CGCATTAAGT 1744
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1498 AGTACAGAAAGGCTGTTCGAGAAATTAACATTCAGTTCTGATCAGAAACGTAAGAAAT 1557
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1745 CGATGCGAGCGACATGTTTGAATTCGCCGATTTCAAGCAACTTGGGTATCATCC 1804
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1558 ACGTGTGCGAGAGCTCATCTGGAATTTGCCGACTTCATGACGAACCATCAACGCTGA 1617
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1805 GAGTGAACGTTAACAAGAGGGTGTTCACCCGTGACCGAAAGCCGAAGGGCGGAGCTC 1864
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1618 GAGTAATCGAAACAAGAGGGATCTTCACTGCGCAGAGACCCCAAACTTCGGCCT 1677
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1865 ATAGTTGAGGGCAAGGTGAGTATGATTA 1898
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1678 TTATTTTGGAGAGATACTGAGAGATTGCCA 1711
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
CVI69266          566 bp  mRNA  linear  EST 09-SEP-2004
LOCUS             rsmxl1.010419.y1.rcf cDNA library of Salvia miltiorrhiza Salvia
DEFINITION        miltiorrhiza cDNA 5', mRNA sequence.
ACCESSION         CVI69266
VERSION           CVI69266.1 GI:51958406
KEYWORDS          EST.
SOURCE            Salvia miltiorrhiza
ORGANISM          Salvia miltiorrhiza
REFERENCE         1 (bases 1 to 566)
AUTHORS           Wang,Z., Yan,Y. and Tian,W.
TITLES            Gene Discovery in Salvia miltiorrhiza Bge. as Revealed by EST
                  Sequencing
JOURNAL           Unpublished (2004)
COMMENT           Contact: Zhezhong Wang, Yaping Yan, Wei Tian
                  Plant Cell Engineering
                  (Wang, Yan) College of Life Sciences, Shaanxi Normal University;
                  (Tian) Hangzhou Genomics Research and Development Institute
                  No. 199 of Chang'an South Road, Xi'an, Shaanxi, 710062, The
                  People's Republic of China
                  Tel: 86-29-85308352
                  Fax: 86-29-85303736
                  Email: zzwang@smu.edu.cn
PCR PRIMERS       FORWARD: M13 R
                  BACKWARD: M13 F
FEATURES          Seq primer: 3'-gtaccagatcgacaa-5' reverse primer M13.
                  Location/Qualifiers
                  1..566
                  /organism="Salvia miltiorrhiza"
                  /mol_type="mRNA"
                  /db_xref="taxon:226208"
                  /sex="hermaphrodite"

```

/!issue type="Whole Youth Plant"
 /dev stage="cultured about 1mon-3mon"
 /clone lib="CDNA Library of Salvia miltiorrhiza"
 /note="Organ: Whole Youth Plant; Vector: plusscript;
 Site 1: EORI; Site 2: XhoII; mRNA isolation used
 PolyAtract (r) mRNA Isolation Systems (PROMEGA, CAT
 No.25200) and plusscript (r) XR CDNA Library Construction
 Kit (Invitrogen, CAT No200455.) to construct the cDNA
 library."

ORIGIN

Query Match 11.0%; Score 208.6; DB 7; Length 566;
 Best Local Similarity 62.6%; Pred. No. 6.4e-54;
 Matches 325; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 1251 AGATGCGATTAAACATTAACCCAGAGGCCCAACAGCCGATTCGTGAGCTCATTCG 1310
 DB 48 AGAGGCACTCAACCGGAACTCGACAGCGCACTTACAGGCCATTAAAGCTGATAGC 107
 QY 1311 CCGAGCAAAAACCCAGCCAGGTGTTCATGTGTCTATTGCCACGAGCCCGCATCTCA 1370
 DB 108 GCGTGACAAAACCCAGCCAGCGTGATGTGAGTATGCCAGAACCGGATACCCG 167
 QY 1371 TGAAGATGAGCTCGGCAATTTCTGAGCCACTGACCAATTTGACTGTCACTTGATCC 1430
 DB 168 TCGGCAAGTGCAGGGAAATATTCGGCCACTGGCGAGCAACCGGTAACTCGACCC 227
 QY 1431 AACTCGCCTATTATTAATTTCTAACTCGGCAAGCGCAACATCAGCTGATCGATCTC 1490
 DB 228 GAGCGCTCCATCACTCGCTCATATGTTCTGAGAGCTTCAACCGATACATCAG 287
 QY 1491 TGATCTGTGATGTCAAGTTGCAATTAATCGGATTTTCGATGTATTTCTCAACAGAGA 1550
 DB 288 CGATCTCTTTGATGTGTGTGCTGCTGAACCGTATTAAGATGTATGTCCAAAGCGCA 347
 QY 1551 CCTTGAAGAGCAGAGGCACTCTTGAAGAAGAGCTGCAGATGAGCAAGAAATTCCA 1610
 DB 348 TTGGAAACGCGAGAGAGTACTGGAAGAAAGCTTGGCTGGAGAGAACTGCA 407
 QY 1611 CAGGCGCATGTCATGACCGAATATGTGTGAGATACCTTCAGAGCCTTCACTTCTCT 1670
 DB 408 TCAGCGCATATCATCAACGAAATACGCGGTGATAGCCGGCTGACATCATGTA 467
 QY 1671 CGAGATGCTTGGAGGAAAGATTCCAGTACAAATGCTAGATGATACATGATGATGTT 1730
 DB 468 CACCGCATGTGAGTGAAGATGATCATGTGATGATGATGATGATGATGATGATGATG 527
 QY 1731 TGATCGCATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1769
 DB 528 TGATCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 566

RESULT 4
 LOCUS CD014092 1995 bp mRNA linear EST 21-OCT-2003
 DEFINITION 90134967 Single gene library Homo sapiens cDNA, mRNA sequence.
 ACCESSION CD014092
 VERSION CD014092.1 GI:3777621
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1995)
 Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R.,
 Au-Young, J. and Stuve, L.L.
 PCR isolation and cloning of novel splice variant mRNAs from known
 drug target genes
 Genomics 83 (4), 566-571 (2004)
 COMMENT
 JOURNAL Incyte Corporation
 3160 Porter Drive, Palo Alto, CA 94304, USA
 Tel: 650 621 8639

Fax: 650 621 8965
 Email: pjin@incyte.com.
 Location/Qualifiers
 FEATURES
 source

1.1995
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone lib="Single gene library"
 /note="Vector: pDrive Cloning Vector; RT-PCR was performed
 using gene-specific primers flanking the open-reading
 frame. PCR products were subcloned into pDrive Cloning
 Vector and sequenced completely using M13 forward and
 reverse primers. Sequencing gaps were closed by
 re-sequencing using primers flanking the gapped areas."

ORIGIN

Query Match 10.8%; Score 205.8; DB 6; Length 1995;
 Best Local Similarity 50.1%; Pred. No. 7.7e-53;
 Matches 688; Conservative 0; Mismatches 607; Indels 78; Gaps 4;

QY 553 ATCAGCAGAGGAGCGGCAAGTCAAGAAATCCAGACTATCAACATGACTTTAACAAC 612
 DB 399 ATGCCATCGGTATCCCAAGGTTACTTTGTCCGAACATATTTGACTTTTCAAC 458
 QY 613 TATGCTGTCTCGCCGATCTATCTGCTTATTTCTGTACCCCAACAATATCCAGAT 672
 DB 459 TAGCTGACCTCAGCGGTCTGTACTTGTGTACAGACACCCACCACTACATGATAC 518
 QY 673 ATTACTGTGTGACAGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 732
 DB 519 ATCACCCTGACCAACAGGCTGAGAGCAAGTGTGGCTGTGATTAATCAAGATCTCTGTC 578
 QY 733 GCGAACACAGACAGGAGGAGATCCAGATCTCAGATGCAAGAGATGAGATTTGTT 792
 DB 579 AAGGCGATTAACCTGTTCAGTGAAGTGCCTTTTGATGACGAAACAAAGTCTTG 638
 QY 793 GCAAGGCTCTGGAGCTCAGGATCTGTCAATTCCTCAGTCAAGCTATG----- 846
 DB 639 GCGAATGGGACTGGGACCAAGGCACTTAAGGTGCCAGGTGACGCTCTGGTGGCG 698
 QY 847 -----CAACTGGCGCCGATTTCTCTACCAACTCCAGTCAACATCGTG 891
 DB 699 TACTGATGACGAAGCGCTGCTGATGTATGATGATGATGATGATGATGATGATGATG 758
 QY 892 GGTTCAGCGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 951
 DB 759 AGTCACTGGGCGCTGTGTGTGATCTTACACACTCCCTGTGGGATCCGACTGTGCT 818
 QY 952 GTTCCGCGGTCAAAATTTTAATAAGAAAGCTTTCTACTTACCGGTTTGGCAAA 1011
 DB 819 GTCAACCAAGACGAGTTCTCATCATGGAACCTTTCTATTTCCAGCGGTGTCAACAG 878
 QY 1012 CATGAAGACAGACAGTACGTGCAAGAGACATGACCCAGATATAGTTCACGATTTTC 1071
 DB 879 CATGAGATGCGGACATCCGAGGAAGGCTTTCATGCGCGCTGTGTGAAGACTTC 938
 QY 1072 CAACTCATGAATGGATTTGAGCAAAATCTTTTGGCACTTCACTAATCTTACGCGGAA 1131
 DB 939 AACCTGCTTCGCTGCTGT 998
 QY 1132 GAGTGCATGATTTGCGAGATGCAATGAAATGATGATGATGATGATGATGATGATGATG 1191
 DB 999 GAAATGATGACAGATGTGTGACCGCTATGGAATGTGTGTGTGTGTGTGTGTGTGTGT 1058
 QY 1192 GGTGGAACATTTGCTGTGATGAGGTGTGTGAGGTGTGTGTGTGTGTGTGTGTGTGTGT 1251
 DB 1059 GGCCTGGCGCT-----GCCG 1073
 QY 1252 GATGCGATTAAGATTAACCAAGAGGCCACAGAGCGGATTCGTGAGCTCATTTGCC 1311
 DB 1074 CAGTTCTTCACCAACGTTTCTGTGATCACCACATGACAGATGAGAAAGTGTGCTT 1133
 QY 1312 CGAGCAAAAACATGCCAGT 1371

[illegible]

/codon_start=1
 /product="Ac2-223"
 /protein_id="AAP6274.1"
 /db_xref="GI:32527745"
 /translation="MKEDIKQSGVEKCYLMDLNSRVLQANERDRESEAKPSRWTAGI
 PAMSPRSRYCVFLGQLCSLALALGQGLFPETSRSEKLVLDGMSPADUSNNR
 LQAPKQWYKQPLREGSPRTLDMVPSSFNQITQSAERLNFGLWVGYREAYLPPQMT
 DTRRIVYLRLNSAHYAIVVWVNGIHVYEHGGHLPFADITKLQVQSGPLTFRRITIA
 NNTLVPTLPPGTVKTKDSMTPKGIQVDDISDFPNVAGLHNSVLAITPTIYID
 IATVTDVDRGLVNWMI SVQSDHDFQLEVRLLDEDKIVARGTNGEQLVPRAHLM
 WPLIMHEHPAYLLELVMTTPPSVSFPLYVQVIRVAATYKSKFLINCKGFYQGVN
 KHEDSDIGRGFWMPPLIKDFENLLRWLGANSFRSHYPSSEVQLCDRVAIVIDE
 RGCVTLPOSFGVSLRHLHLEWDELVRDKNHDAVVMMSVANEPSVSLKAGYQFPI
 IVQGLKTSDSPTFGTVALVNYHAPVDVYCNLSVSYNDYGLLEVIQLOLDSQFEN
 WYMQGKPTIIQSEYGVADVASGLHSDPRMHSSEYQVALLLENYHLILBEKREYIYGEI
 IANFADMTQNSPLRYVGNKKGITFRQRNQMDSVAILRERVYRANSTRGILIQVQNS
 IVNDEKEVQHLICGADVDATLLQHKVQNVCGDSVIAAGALICSYCTNRRAHELCKLST
 VPSQDAGAEQGGQLSIQATGMVNTINKACSHFQVRLQKTHVHNSIAVFKPDPLRSR
 CLLDISAGSA"

Query Match	9.9%;	Score 188.4;	DB 3;	Length 2473;
Best Local Similarity	49.0%;	Pred. No. 2,78-47;		
Matches 833;	Conservative 0;	Mismatches 756;	Indels 111;	Gaps 8;
QY	247	CCAGTCCCGGCTCTTACAAAGACATCTTCATGACCGGAGATTCAAGACCATGTGGGA	306	
Db	413	CCGGTCCCTTCAGAGCTTCAATGACATCAACCAAGAACAGAGCTTCGGAATCTTATGGC	472	
QY	307	TGGGTTTACATATCAGCGTAGTCATGTGTCCCAAGCGTGTCTCAGGA-----G	357	
Db	473	TGGGTGTGTATGAACCGGAAGCAGTCTTCCACAGCGATGAGACCCAGAACACGACAG	532	
QY	358	CGATATCTCGTGGCAGCCGATCCGCTACGCAACCATGATGTGCATTTATGTCAACACCG	417	
Db	533	AGAGTGTGTATGAATCAACAGCGCCCATTTACTATGCAGTGTGTGGTGAATGGGATT	592	
QY	418	CTTGTGTCCGAGCATGTGGCGGCTATATACCTTTGAAGGAGCGTCACTGAATTAGTC	477	
Db	593	CATGTGTGTGAACATGAGGAGGTCACCTCCCTTTGAGCGTGACATCACCAACTGTGTC	652	
QY	478	-----GCCCCCGGAGGAATTTTCGTTTGAACATTTGTGTGCAACAGAGCTTAACCAT	531	
Db	653	CAGAGTGGGCCCTTGACCACTTCCGGGTACCATTCGCATCAACACACTGACCCCT	712	
QY	532	GAGACTATCCCAACCTGGAAAAATCA-----CGACAGGGAACGCACTGGC	576	
Db	713	TATATCCCTTCCACCGGAGACCATTTGTCTACAGACTATCTTTCATGTATCCAAAGGT	772	
QY	577	AAGAGATTCAGACCTTATCAACTGACTTTTCAACTATGTGTGTCTCCGCCGATCTATC	636	
Db	773	TACTTCGTCCAGGACATATAGCTTCGACTTCTTCAACTATGCGGGGTGCAACGGTCTGTG	832	
QY	637	TGGCTTATTTCTGTACCCCGACCAACATATCCAGGATTTTACTGTGTGTACAGATGTTGAT	696	
Db	833	GTCTGTACACCAACCCCTTACACCTTATATGATGATATCACTGTGACCACTGAGGTGAC	892	
QY	697	GGTGAACAATGCTCTGATTAATCTACGAGGTGGAAGTGGCGAACGACGAGCGGGGACATC	756	
Db	893	CGGAGCCTTGGCGCTGTGTAACCTACTGTGATTTCTGTCCAGGGCAGTGAACATTTCCAGCTA	952	
QY	757	CAGATCTCAGTGTATCGACGAGATGAGACTATTTGTTCAAAGGCTCGGAGACTCAGGGT	816	
Db	953	GAAAGTGTCTTCTGTGATGAGATGGAATAAATTGTGCGCGTGGAACAGGGAATGAGGGGT	1012	
QY	817	ACTGTGACAAATTCCTTCAGTCAAGCTATATGGCAACTGTGGCGCGGATATCTCTACCACTC	876	
Db	1013	CAACTTAAGGTGCGCGTGCACCTCTGTGTGGCTTTACTGTATGATATGAGCACTCAGCC	1072	
QY	877	CAGGTCAACATCGTGGGTTCTAGC-----GGCATGTATGTCGACACTCAACAT	924	
Db	1073	TACCTGTACTCTTTGGAGGTGACGATGACAAACACTGATGTCTGTGTCTGACTTTCTATACC	1132	

```

QY 925 TTGGCTACGGGCGTGCCTACTCTCAAGTTGCCGGTCAATTTTAATAATGAAG 984
Db 1133 CTCCTCTGTGGGATTTGGAACAGTGGTGTACAAAGAGAGTTCTCTAATAATGGAG 1192
QY 985 CCTTCTACTTACCGGTTTGGCAACATGAAGACAGCATGTCGCGCAAGAGACAT 1044
Db 1193 CCTTCTACTTCCAAAGCGGTCAACAGCATGAGATTCAGATATCCAGAGGAGAGCTTC 1252
QY 1045 GACCCAGCATACATGCTGCATGATTTCCAACTCATGAATAATGATGGAGCAATCTTTT 1104
Db 1253 GACTGGCTCTGTGTATTAAGATTTCAACCTCTCCGTTGGCTGGGGCAATCTCTT 1312
QY 1105 CGGACTTTCACATCTCTTACCGCGAAGGTCATGATTTTGCAGATCGAATGGAAT 1164
Db 1313 CGTACCGACCATCTCTCTACTCTCGAGAGGATCTACGCTGTGACCGATATGGAAT 1372
QY 1165 GTCTGATGCATGAATAACCTGCGGTGTGTGATTCGATCTTATGAGGCGATCTGAG 1224
Db 1373 GTGGTATGATGAGTGTCCGCGTGTGGCATCTGCTGCCCAAGTT----- 1421
QY 1225 AGTGTGCCCAACAAACATTTTACGCCAGATGCGATTAACGATTAACCCAGAGGCCAC 1284
Db 1422 -----TTGCAACGTGTCTCTTCCGACAC 1447
QY 1285 AAGCAGGCGATTCGAGCTCATTTGCCGAGACAAACCATGCCAGTGTTCATGTG 1344
Db 1448 CTAGAGGTGATGACAGAGCTGTGCGAGAGACAAATACTCCGCGGTGTATGTGG 1507
QY 1345 TCTATTCGCAACGAGCCCGCATCTCATGAATGAGAGCTCGCAATCTTGCAGCCACTG 1404
Db 1508 TCTGTGCGCAATGAGCTGTCTCTCTGTGAACCTCGCGATTTACTTCAAG----- 1561
QY 1405 ACCAATTGATCTCGCATCTTGCATCTGCACTGCCCTATTACATTTGCTAACGTGCGACG 1464
Db 1562 -----CTATGTTCTTCAAGGCTTGAAGACATTCAGATTTCCCAAGAACTGGA-- 1609
QY 1465 GCGACATATCAGCTGATCGGATCTGATCTGTTTGAATGTCAGTATTAATTCGATAT 1524
Db 1610 GTTACAGCTCTTGAATACATCAATCCCGTACGTGACGTGATTTGTGAACAGTTAC 1669
QY 1525 TTTCGATGATTTCTCAACAGAGACCTTGAAGAGACAGGACGCTTTGAAAAGAG 1584
Db 1670 TTATCTGATATCATGATCAAGGACATCTGAGAGTGTATGAGTGCAGCTGATCAG 1729
QY 1585 CTGATGATGAGCAAGAAATTCACAGGCGCATTCGATGACCGCAATTTGTGTGAGAT 1644
Db 1730 TTGAGAACTGTATTAAGATGTACAGAGCCATTTATCCAGAGCGAGATGAGAGAC 1789
QY 1645 ACCCTTGAAGCTTCACTCTATCTGCACTGCTGAGCGAGAGATTCAGATACAA 1704
Db 1790 GCCGCTCGGGGCTTCAATGAGATCCACCTCGCATGTCAGTGAAGATACAGACAGCT 1849
QY 1705 ATGCTAGACATGATACATGAGTGTGAT-----CGCATTAATGATGAGCGAG 1758
Db 1850 CTCTGAGATATTAATTTGATCTCTGATGAGAGAAAGAAATATATCTCATCTCGAGAG 1909
QY 1759 CATGTTGATCTTCCGCAATTCAGACCAATTTGGTATCATCTGAGTGAAGCGTAC 1818
Db 1910 CTATCTGAGATTTTGTGATCTTCAAGAGAACATCACTACACTAGAGTACAGAGAAC 1969
QY 1819 AAGAAGGATGTTTTCACCGCTGACCGAAAGCCAAAGCGCGCACTCATATGTTGAGGCA 1878
Db 1970 AAGAAGGATGTTTTCACCGCTGACCGAAAGCCAAAGCGCGCACTCATATGTTGAGGCA 1878
QY 1879 AGGTGATGATGATTAATA 1898
Db 2030 AGATACTGAGGATTCGCA 2049

```

RESULT 6
 CR593823
 LOCUS CR593823 1377 bp mRNA linear HTC 21-JUL-2004
 DEFINITION full-length cDNA clone CS0DL001YM21 of B cells (Ramos cell line)

```

ACCESSION .Cot 25-normalized of Homo sapiens (human).
VERSION CR593823.1 GI:50474630
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1377)
JOURNAL Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
REMARK Full-length cDNA libraries and normalization
Contact : Feng Liang Email : fliang@life.technet.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
REFERENCE Faraday Avenue
AUTHORS 2 (bases 1 to 1377)
TITLE Direct Submission
JOURNAL Genoscope.
COMMENT BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo (dt) primer. Five prime
end enriched, double-strand cDNA was digested with NotI and cloned
into the NotI and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by life technologies, a
division of invitrogen.
FEATURES
Location/Qualifiers
source
1..1377
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL001YM21"
/tissue_type="B cells (Ramos cell line) Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 9.6%; Score 183.2; DB 3; Length 1377;
Best Local Similarity 51.2%; Pred. No. 9.8e-46;
Matches 551; Conservative 0; Mismatches 468; Indels 57; Gaps 3;
QY 829 CCTGATGCAAGCTATGAGCAACCTGGCGCGCATATCTTACCAATCCAGGTACATC 888
Db 145 CCGTACGTGATGACAGAAACGCTGCTTCTGATTCATTTGAGAGTGCAGTCACTGCA 204
QY 889 GTGGGTTTACGCGGAGTATGATGACACCTTACATTTTGGCTACGCGGCTGATCTC 948
Db 205 CAGAGTCACTGGGGCTGTGTCTGATCTTACACATCTCCCTGGGAGATCCGACTGTG 264
QY 949 AAGTTGCGGGGTCAATTTCTTAATAATGAAGCTTTCTACTTACCGGTTTGGC 1008
Db 265 GCTGTACACCAAGACCAGTCTCTCATCATATGGAACCTTTCTATTTCCACGGTGTAC 324
QY 1009 AATCATGAAGACAGAGATGAGTGAAGAGACATGAGCCAGCATATCATGTTCAGAT 1068
Db 325 AAGATATGATGAGTGGAGCATTCAGAGAGAGGCTTGCAGCTGCGCTGTGTAAGAGAC 384
QY 1069 TTCAACTCATGAATGATGAGCAATTTCTTTTTCGACTTCACTATCTTACGCG 1128
Db 385 TTCAACTGCTGCTGCTGCTTGTGCGCAACGCTTTCGATACAGCCACTACCCCTATGCA 444
QY 1129 GAAGAGTCAATGATTTGCGAGATGCAATGCAATGATGCTGTGATGATGAACACCTGCC 1188
Db 445 GAGGAGTGAATGAGATGCTGACCCCTATGAGATGCTGATCATGATGATGCTCCGCGC 504
QY 1189 GTTGTCTGAACATTTGCTGATGAGCGATCTGAGAGTGTGCCCAACATTTAGC 1248
Db 505 GTGGGCTGGCGT-----G 519
QY 1249 CCAATGCAATTAAGATTAACCAAGAGGCCCAAGACAGCGATTCGTGAGCTATT 1308
Db 520 CCGCATTTCTTCAACACGTTTCTGTGATCAACACATGAGTATGGAAGAAATGTGTG 579
QY 1309 GCCGAGACAAACCATGCGAGTGTGATGATGTGTCTATTGTCACAGAGCCGCACT 1368

```



```

DB      580 CGTAGGAGCAAGAACACCCCGCGTGTGATGTGTGTGCGCAACGAGCTGTGCTCC 639
QY      1369 CATGAAGATGAGAGCTGCGAATCTTCGAGCCACTGACCAATTGACTGTCACTGAT 1428
DB      640 CACCTTAGAATCTGTGGTACTACTTGAAGATGGTATGCTTCAACCAATCTTGGAC 699
QY      1429 CCAACTGCGCCTTATTCATTTTGCTAACTGTGCGCAGCGCGCATATAGGTGATCGGATC 1488
DB      700 CCTCTCCGGGCTGTGACCTTT-----GTGAGCACTTACTTACTTACAGACGAAAGGG 753
QY      1489 TCTGATCTGTGATGTGATGTCATGTCATTAATCGGTATTTGCGATGATTTCTCAACAGCA 1548
DB      754 GCTCCGTATGTGATGTGATGTCGTGTAACAGCTACTACTTGTGATACGACTACGCG 813
QY      1549 GACCTTGAGAGAGCAGAGCAGCTCTTGAAAGAGAGCTGATGATGCGCAAGAAATTC 1608
DB      814 CACTGTGAGTGTGATTCAGCTGACGCTGGCCACCCAGTTTGAAGACTGTATAGAAATAT 873
QY      1609 CACAGGCGATGTCATGACCGAATATGGTGAGATACCTTGCAGGCGCTTCACTATC 1668
DB      874 CAGAGCCCATTTATTTAGAGCGAATGATGACAGAAACGATTCAGAGGTTTCAACAGAT 933
QY      1669 CTCGAGCTGCTTGGAGCGAAGATTCAGATCAAAATGCTAGACATGATCCATCGAGTG 1728
DB      934 CCACCTGTGATGTTCACTGAAGATGACGAAAGTCTGTAGAGACATGATCGGGT 993
QY      1729 TTGTAT-----CGCATTTGATGATGACGAGCGAGCATGTTTGAATTCGCCGATTC 1782
DB      994 CTGGATCAAAAACGCGAATAATCGTGTGGAGACTCATTTTGAATTTTGGCGATTC 1053
QY      1783 CAGACCACTTGGGTATCATCCGAGTACGCTTAACAAGAGGCTTTTCAACCGGTAC 1842
DB      1054 ATGACTGACATGTCACCGACGAGAGTCTGGGAGATTAAGGAGATCTTCACTCGGAC 1113
QY      1843 CGAAAGCCAAAGCGCGACTCATAGTTTGAGGCGAAGGTGACATGATTTGATA 1898
DB      1114 AACAACAACAAAAGTCAGCGTCTTTTGGAGAGATACGAAAGATTCGCA 1169

RESULT 7
CD014093      1853 bp      mRNA      linear      EST 21-OCT-2003
LOCUS      90135266 Single gene library Homo sapiens cDNA, mRNA sequence.
DEFINITION      CD014093
ACCESSION      CD014093
VERSION      CD014093.1 GI:37777622
KEYWORDS
SOURCE
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1853)
Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R.,
Au-Young, J., and Stuve, L.L.
PCR isolation and cloning of novel splice variant mRNAs from known
drug target genes
Genomics 83 (4), 566-571 (2004)
CONTACT: Jin, P.
Incyte Corporation
3160 Porter Drive, Palo Alto, CA 94304, USA
Tel: 650 621 8639
Fax: 650 621 8965
Email: pj@incyte.com.
Location/Qualifiers
1..1853
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_11b="Single gene library"
/notes="Vector: pDrive Cloning Vector; RT-PCR was performed
using gene-specific primers flanking the open-reading
frame. PCR products were subcloned into pDrive Cloning
Vector and sequenced completely using M13 forward and

```

```

ORIGIN
Query Match      9.6%; Score 183.2; DB 6; Length 1853;
Best Local Similarity 51.2%; Pred. No. 1.1e-45;
Matches 551; Conservative 0; Mismatches 468; Indels 57; Gaps 3;

reverse primers. Sequencing gaps were closed by
re-sequencing using primers flanking the gapped areas."

QY      829 CCTCAGTCACTATGCGAATGCGCGCGCATATCTTACCAACTCCAGGTCACATC 888
DB      554 CCGTACTATATCAGCAAGCCCTGCTATCTGTATTCATGAGAGTGAAGCTGATCGCA 613
QY      889 GTGGGTTCTAGCGCGCATGTAGTCACACTTCAATTTGGCTACGCGCGTGTACTGTTC 948
DB      614 CAGAGCTCACTGGGGCTGTGTCTGACTTTCACACTCTCCGTGGGATTCGCACTGG 673
QY      949 AAGGTTCCGGGTCACAATTTCTTAATTAATGAAAGCCCTTCTACTTACCGGTTTGGC 1008
DB      674 GCTGTCAACAAGAGCCAGTCTCTCATTAATGAGAAACCTTCTATTTCCACGGTCAAC 733
QY      1009 AAACATGAAGACACAGCACTAGTGGCAAGACATGACCCAGCATACATGTGTCACAT 1068
DB      734 AAGCATAGAGATGCGGACATCCGAGGAGGCGCTTGCATGCGCCGCTGCTGTGAAGAC 793
QY      1069 TTCCAACTCATGAATGATGATGAGCAATTTCTTTTGGACTTTCACACTATCTTACGGG 1128
DB      794 TTCAACTGTCTTGTGCTGCTGTGTGTGCAACGCTTTCCGATCAGACCATACCTTATGCA 853
QY      1129 GAAGAGTCATGATTTTCGACATGCAATGCAATTTGCTGATGATGATGAAACCTGTGC 1188
DB      854 GAGAAATGATCAGATGTGTGACCGCTATGGGATGTGTGATGATGATGTCCGGC 913
QY      1189 GTTGTCTGAATTCCTTGAATGGCGCTTATGAGAGTGTGCCCAAACTTTAG 1248
DB      914 GTGGGCTTGGCGCT-----G 928
QY      1249 CCAGATGCGATTAAGATTAACCAAGAGCGCCACAAAGAGCGATTTGTGAGTCATT 1308
DB      929 CCGGACTTTTCAACAACGTTTCTTGTGATACCACTAGCTGATGAGAAAGTGTG 988
QY      1309 GCCCGAGCAAAAACCATGCGAGTGTGTATGTGTCTATTGCAACGAGCCCGCATCT 1368
DB      989 CGTAGGAGCAAGAACACCCCGGTCGTGATGTGTGCGCAACGAGCTGTGCTC 1048
QY      1369 CATGAAGATGAGAGCTGCGAATCTTCGAGCCACTGACCAATTTGACTGTCACTGAT 1428
DB      1049 CACTTAGAATCTGTGCTACTACTTGAAGATGTGATGCTCACCAAAATCTTGGAC 1108
QY      1429 CCAACTGCGCCTTATTAATTTGCTAACTGTGCGCAGCGCAATTCAGCTGATCGGATC 1488
DB      1109 CCTCTCCGGGCTGTGACCTTT-----GTGAGCACTTACTTACTTACAGACGAAAGGG 1162
QY      1489 TCTGATCTGTGATGTGATGTCATTAATCGGTATTTGCGATGATTTCTCAACAGCA 1548
DB      1163 GCTCCGTATGTGATGTGATGTCGTTTGAACAGCTACTACTTGTGATACGACTACGCG 1222
QY      1549 GACCTTGAGAGAGCAGAGCAGCTCTTGAAAGAGAGCTGATGATGCGCAAGAAATTC 1608
DB      1223 CACTGTGAGTGTATTCAGCTGACGCTGGCCACCGATTGAAACTGTGTATAGAAATAT 1282
QY      1609 CACAGGCGATGTCATGACCGAATATGGTGAAGATACCTTGCAGGCGCTTCACTATC 1668
DB      1283 CAGAGCCCATTTATTCAGAGCAGATGAGCAAGAAACATTTGACGGGTTTCAACAGAT 1342
QY      1669 CTCGAGCTGCTTGGAGCGAAGATTCCAATGATCAAAATGCTAGACATGATCACTGAGTG 1728
DB      1343 CCACCTGTGATGTTCACTGAAGATGACAGAAAGTCTGTAGAGAGTACCATCTGGGT 1402
QY      1729 TTGTAT-----CGCATTTGATGATGACGAGCGAGCATGTTTGAATTCGCCGATTC 1782
DB      1403 CTGGATCAAAAACGCGAATAATCGTGTGGAGGCTCATTTTGAATTTTCCGATTC 1462
QY      1783 CAGACCAACTTGGTATCATCTGAGTACGCTTAACAAGAGGCTTTTCAACCGGTAC 1842

```

Db 1463 ATACTGACAGACGACCGACGAGAGTGCAGGGAATTAAGGAGATCTTCCTGGCAG 1522

Qy 1843 CGAAGCGCAAGCGGCGACACTACTAGTTGAGGGCAAGGTGACATGATTGATTA 1898

Db 1523 AGCAACCAAAAGTACGCGCTTCTTTGCGAGAGATCTGGAAGATTGCCAA 1578

RESULT 8
BX745933/c
LOCUS BX745933
DEFINITION 857 bp mRNA linear EST 18-NOV-2003
MENA sequence.
BX745933
VERSION BX745933.1 GI:38418673
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE
AUTHORS Croxson, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.
TITLE Sanger Xenopus tropicalis EST project 2001 (11_2003)
JOURNAL Unpublished (2003)
COMMENT Contact: Croxson MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: Tgas066d19.g1kx77
Sequencing primer: T7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron W. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-blue.

FEATURES

location/Qualifiers
1..857
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="Tgas066d19"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_id="XGC-gastrula"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 9.5%; Score 180.4; DB 5; Length 857;
Best Local Similarity 54.4%; Pred. No. 6.4e-45;
Matches 444; Conservative 0; Mismatches 321; Indels 51; Gaps 2;

Qy 907 GTATGACACCCACCAATTTGGCTACGGCGCGCTGCTCAAGTTCGGGTCACAA 966

Db 789 GTGGAAGACAAATTAATTTGGCTGCTGGAATTAAGACGTGCACGTCTCTGGAGACCAA 730

Qy 967 TTCTTAATAATGAAGCCTTCTACTTACCGGTTTGGCAACATGAAGACACACA 1026

Db 729 TTCTTCATCAATGAAGAACATTTACTTCATGCTGCACAAACATGAAGACATATAT 670

Qy 1027 GTACGCGCAAGGACATGACCCAGATACATGCTTCACGATTTGCAACTCATGAATGG 1086

Db 669 GTCAAGGAAAGGACATGACCTGCTCACTAATTTGAAGATTTTAATCTGTTGAAGTGG 610

Qy 1087 ATTGAGCAAAATTTCTTTGCGACTTCACACTATCTTAACGCGGAAGGTATGATTTTC 1146

Db 609 CTGTGCTACTCTCTCCGACACAGCCATTAATCTTATGACAGAAATCATGACCTT 550

Qy 1147 GCAGATGGAATGGAATTTGTCGATGATGAAGAAACCTCCGCTGTGTCTGAACATTTGCC 1206

Db 549 TGTGATTAATATGACATTTGTTGATGATGAATGCTCCGAGGTGATTA----- 497

Qy 1207 TTGATGGCGGTATCTGAGAGTGTGCCCAAAACATTTACCCAGATGCGATTAAGCT 1266

Db 496 -----ATACCCGAAAGTTTGGAAAC 475

Qy 1267 AAAACCAAGAGGCCACACAGAGGATTCGTGACCTCATTTGCCGACAAACCAAT 1326

Db 474 CAATCTTTAAACACCATTTAAATAGTTATGAGAGGATTTGCGAGGATTAACCAACCG 415

Qy 1327 GCCAGTGTGATGATGCTGCTATTTGCAACGAGCCGCGATCATGAAGATGAGCTGCG 1386

Db 414 CTTCCGTTGATGATGCTGCTGCTGCAATGACAGATCCAGCTCCCTGTGCTGGG 355

Qy 1387 GAATACTTGAGCCACATGACCAATTTGACTGCTCACTTGAATCCAACTGCGCTTAATTA 1446

Db 354 TATTTATTTAAACGATGATTTGTTACATTAAGCAACTTGACCCACCCGACAGTGACA 295

Qy 1447 TTGCTTAAGCTGCGACCGGACATATCATGCTGATGCGATCTGATCTTTGATGTC 1506

Db 294 TATGTTTCAATGCTTAATATGAACATGACCAAGTGCACCTTATG-----TGATGTA 241

Qy 1507 ACTTGCAATTAATCGTATTTGGATGATGATTTCTCAACAGAGACCTTGAGAAAGCAGAG 1566

Db 240 ATTTGTGTAACAGTACTTTTCTTGTATCATATGCGCGGACCTGGAAGTATCCAA 181

Qy 1567 GCAGCTCTTGAAGAGAGCTCATGAGTGAAGAAATTCACAGCCGATGCTATG 1626

Db 180 CTCCAATCAATGATCACTTTGACAAATGATATGAAGATCAAAAGCCTAATGATACAG 121

Qy 1627 ACCGAATTTGTCGATGATCCCTTGACGCGCTTCACTCTATCTCGGACTGCTTGAGC 1686

Db 120 AGTGAATATGAGAGATCAATTCAGAGGTTTCAAGTACGCCCATTAATGTTCACT 61

Qy 1687 GAAGAGTTCCAAATGCTGACATGATACAT 1722

Db 60 GAAAGATTCAGAGAGTGTGTGAAGAACTACCAT 25

RESULT 9
LOCUS CL679241
DEFINITION 811 bp DNA linear GSS 09-JUL-2004
PRI0125C_C10_2 - PRI0125C.BR (811) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic
survey sequence.
CL679241
VERSION CL679241.1 GI:50185841
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Nedoplogasteridae; Pristionchus.
1 (bases 1 to 811)
Stinivaasen, J., Otto, G.W., Kahlow, U., Geisler, R. and Sommer, R.J.
Appab: an Acedb database for the nematode satellite organism
Pristionchus pacificus
Nucleic Acids Res. 32 (1), D421-D422 (2004)
Contact: Sommer, R.J.
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@uebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.
Location/Qualifiers

FEATURES

source
 1. .811
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strain="California"
 /db_xref="taxon:54126"
 /clone_lib="Mixed stage fosmid library of P. pacificus
 var. California"
 /note="Vector: pBfi105-5 Fosmid vector"

Query Match 9.4%; Score 178.8; DB 9; Length 811;
 Best Local Similarity 53.8%; Pred. No. 2e-44;
 Matches 444; Conservative 0; Mismatches 367; Indels 15; Gaps 3;

QY 289 ATTCAAGACCATGAGATGAGGTTTACTATCAGAGAGGTCATTTGCCCAAGGCTGG 348
 DB 1 ATTGTAATTTATTTGGCAACGCTGCTATCAGCGGCAAGCTTTATACGAAAGGTTGG 60
 QY 349 TCTCAGAGCGATATCTGTCGAGCGGATCCGCTACAGCAGCATGGTCATCTATGTC 408
 DB 61 GGAAGCGAGGATTCGTCGTCGCTTCGATCGGCTACCTATTAAGCAAGTGCGGTC 120
 QY 409 AACACCCGCTTGTGCGAGCATGTGGCGGCTATACCTTTGAAGCGACGTCAT 468
 DB 121 AATTAATCAGAAATGATGAGCATCAGGGCGGCTATACGCAATTTGAAGCGATGTCAG 180
 QY 469 GAATTAATGCGCCCGGAGAAATTCGCTTACCATTTGGTCAACAGAGCTTAC 528
 DB 181 CCGTATGTTATTCGCGGAAAGTGAAGTATACCGTTTGTGAAACAGAGCTGAC 240
 QY 529 CATGAGCATATCCACCTGGAAAAATACACAGAGGAAACGCACTGGCAAGAAATCCAG 588
 DB 241 TGGCAGACTATCCGCGGGAATGATTAACG---ACGAAACGCAAGAAAGGACG 297
 QY 589 ACCTATCAACATGACTTTTACAATGCTGTGCTGCGCCGCTATCTGGCTTTATCT 648
 DB 298 TCTTACTTCTCATATTTCTTAACTACGCGGAGATCCATGCGACGCTTAATGCTCTAC 357
 QY 649 GATACCCAGAAATATCCAGATTTACTGTGTTTACAGATGTTGATGTTGACATGCT 708
 DB 358 AGCGCAACACTGGGTGAGCATATACCGTGTGACGATGCTCGCAAGCTGTGAC 417
 QY 709 CTGATTAATACAGAGTTCGAAAGTGGCAACAGACAGCGGAGAGATCCAGATCTCAGT 768
 DB 418 C-----ACGCGTCTGTGATCTGGCAGGTGTGTGCAATGATGATCAGCGTTGAAC 471
 QY 769 ATCGACGAGATGAGCATATTTGTGCAAAAGCTCGGAGCTCAGGATCTGACAAAT 828
 DB 472 CGTATGCGGATCAACAGAGTGTGCACTGCAAGGCAACAGCGGACTTTGCAAGT 531
 QY 829 CCTCATGTAAGTATGAGCACTGTGCGCGCATCTCTACCAACTCCAGGTCACATC 888
 DB 532 GTGATATCCGACCTGTGGCAATGGGTGAAGTTATCTTATGAACTGTGTCACAGCC 591
 QY 889 GTGGGTTCTAGCGCGCATGTGACACCTTACATTTGGCTAGGGCGGTGCTACTGTC 948
 DB 592 AAAAGCAGACAGAGT-----GATATCTACCGCTGTGCGCTCGGATCTCGGTGAG 645
 QY 949 AAGGTTCGCGGTACAAATTTCTTAATATGAAAGCTTTTCTTAACTCGGTTTGGC 1008
 DB 646 GCAATGAAAGGCGCAAGTCTGATCAACCAAAACGTTCTTAACTTCTGCTTTGGC 705
 QY 1009 AAACATGAAGACACAGAGTGTGCAAGAGCAATGACCCAGCATACATGTTGACAT 1068
 DB 706 CGTCAATGAAGTGTGCGGCGCAAGATTCGATTAAGTGTGAGTGTGACAT 765
 QY 1069 TTCAACTCATGAAATGATTTGAGCAAAATCTTTTGGAGCTTAC 1114
 DB 766 CACGATTAATGACTGATGGGCAACTCTTACGATCTCTGCG 811

RESULT 10
 CL463714/c

LOCUS CL463714 908 bp DNA linear GSS 31-MAR-2004
 DEFINITION SAIL_119_H09.v1 SAIL Collection Arabidopsis thaliana genomic clone
 SAIL_119_H09.v1, genomic survey sequence.
 ACCESSION CL463714
 VERSION CL463714.1 GI:45866619
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE
 1 (bases 1 to 908)
 Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
 Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
 Bullis,D., Snell,J., Miguel,T., Hutchinson,D., Kimmery,B.,
 Mitzel,T., Katergiri,F., Glazebrook,J., Law,M. and Goff,S.A.
 A high-throughput Arabidopsis reverse genetics system
 Plant Cell 14 (12), 2985-2994 (2002)

COMMENT

Applied Trait Genetics
 Syngenta Biotechnology Inc.
 3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
 Email: allen.sessions@syngenta.com
 ABRC Stock Number CS805872; T-DNA left border flanking sequences of
 Syngenta Arabidopsis Insertion Library (SAIL) lines are available
 through the Arabidopsis Biological Resource Center (ABRC).
 Sequences represent a pool of amplified genomic regions and not
 single contiguous sequences.
 Class: TDNA tagged.
 Location/Qualifiers

FEATURES

source 1..908
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Columbia"
 /db_xref="taxon:3702"
 /clone_lib="SAIL_119_H09.v1"
 /note="T-DNA left border sequences were isolated using a
 modified TAIL-PCR strategy"

ORIGIN

Query Match 9.2%; Score 175.4; DB 9; Length 908;
 Best Local Similarity 64.3%; Pred. No. 2.5e-43;
 Matches 263; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1480 GATCGATCTGATCTGTTGATGTCAGTTGATTAATCGATTTTCGATGTAATCT 1539
 DB 715 GATACATCAGGATCTCTTGTATGTCGTGCTGAACGTTATACGATGTAATGTC 656
 QY 1540 CAACACGAGACCTTGAGAAAGCAGAGCAGCTTTGAAAAGAGCTGATGATGCGCA 1599
 DB 655 CAAGGCGGATTTGAAACGAGAGAGTACAGGAAAGAACTTCTGCGCTTGCGAG 596
 QY 1600 GAGAAATTCACAGCGCGATCGTCAATGACCGAATATGTCAGATACCTTGCAGGCTT 1659
 DB 595 GAGAACTGATCAGCGGATTTATCATCACCAATACGCGTGATACGTTAGCGGCTG 536
 QY 1660 CACTTATCTCGGATCGCTTGTGAGCGAAGTTCACATGACAAATGCTAGACATGTAC 1719
 DB 535 CACTGAATGATCTCGACATGTGAGTGAAGTATCAATGTCATGCTGATATGTAT 476
 QY 1720 CATCAGTGTGTTGATGCAATGATGATGATGATGATGATGATGATGATGATGAT 1779
 DB 475 CACCGGCTTTGATGCTCAGCGCGCTGTCGTGCGTGAACAGATATGAAATTTCCGCA 416
 QY 1780 TTCCAGAACCACTGGGATTCATCCAGATGACGATGACAAAGAGGATTTTACCCTG 1839
 DB 415 TTTCGACCTCGCAGAGCATATTTGCGGCTGGCGGATGACAAAGAGGATCTTCACTCC 356
 QY 1840 GACCGAAACCAAGCGGACCTCATTAATTTGAGGCGAAGTGAATA 1888

Db 355 GACCGCAAGCGGCTTTTGTGAGCAGACAGACATA 307

RESULT 11
CD503076 1124 bp mRNA linear EST 12-JUN-2003
LOCUS CDA60-C07.5, mRNA sequence.
DEFINITION CDA60-C07.5, mRNA sequence.
ACCESSION CD503076
VERSION CD503076.1 GI:31432977
KEYWORDS EST.
SOURCE Gasterosteus aculeatus (three spined stickleback)
ORGANISM Gasterosteus aculeatus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neotelostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
1 (bases 1 to 1124)
Kingsley, D.M., Petchel, C., Balabhadra, S., Grimwood, J., Dickson, M.,
Schmutz, J., and Myers, R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
Contact: Kingsley, DM
HHMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5934
Fax: 650 725 7739
Email: kingsley@cmgm.stanford.edu
Plate: 60
High quality sequence stop: 782.
Location/Qualifiers
1..1124
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA60-C07"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_id="SHGC-CDA"
/note="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI
(5' adaptor); Site 2: XhoI (3' linker primer); The mixed
organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dT sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."

ORIGIN

Query Match 8.7%; Score 166.2; DB 6; Length 1124;
Best Local Similarity 52.7%; Pred. No. 2.2e-40;
Matches 500; Conservative 0; Mismatches 388; Indels 61; Gaps 4;

QY 912 CGACACCTCAATTTGGCTACGGCGCTGCTACTGTCAGAGTTCGGGGTACAAATTTT 971
Db 73 CGACGTCGTACACTCTACAGTCGGGATCGGACGGTTCACGTTACAGACCCAGTTCTT 132
QY 972 AATTAATGGAAGCCTTTCTACTTTTACCGGTTTGGCAACATGGAAGACAGAGAGTACG 1031
Db 133 CATCAACAAAGAGCCCTTTCTACTCTCCAGAGTAATTAACACGAGACTCTGATATTGG 192
QY 1032 TGGCAAAAGACATGACCCAGCATATGATGTTTCCAGCATTCATGAAGATGATTTGG 1091
Db 193 AGGCAAAAGGCTGAGCTGGCCCTCATGTGTAAGACTTTTATTGAAAGTGTGGG 252

QY 1092 AGCAAAATTTCTTTGAGACTTCACACTATCTTAAGGGGAAGAGTATGATTTCCGAGA 1151
Db 253 GGCCAACTCGTTCCGACACCGACCACTACCTTTAGCAGAGAGATCTCAATGTGTGA 312
QY 1152 TCGAATGGAATTTGTCGTATGATGAAACACCTGCGGTGTTGATGACATGCTTGAT 1211
Db 313 CCGCATGAGCATGTGTGTATGACGATGCGCGGCGTGGG----- 354
QY 1212 GGGCGTATCGAAGTGTGCCCAAACTTTAGCCAGATGCGATTACGATTAAC 1271
Db 355 -----CATTAAGACTTTGCGAGTTTGGAAACCC 385
QY 1272 CCAAGAGCC--CACAAGAGCGGATTCGTGAGCTCATTTCCCGAGACAAACCATGCC 1329
Db 386 TCCTTAACCATCACTGTGTGTCATGACGAGCTGTGATCGTGGGACAAAGAACCATCCC 445
QY 1330 AGTGTGATGATGTGTCTTTTGGCAACGAGCCCGCATTCATGAAGATGAGCTCGGAA 1389
Db 446 TCTGTGATGATGTGTGTCAGTGGCAATGAGCCGCTGCAAGATGCTCTGTGATTA 505
QY 1390 TACTTCAGACCACTGACCAATTTGACTGTCATCACTGATTCGACCTGCTTATTAATTT 1449
Db 506 TATTTCAAAACCTTGATTAACATACCAAGATTTGATTCACCCGCGCTGCACTTT 565
QY 1450 GCTAAGTGTGCGACGCGCATATGACGTGATGATCTGTGATCTGTTGATGTCAGT 1509
Db 566 ATCAGAGACG-----TACTATGCCAGGATTAAGGGGCTCTCCCTAGTGAATC 619
QY 1510 TGCATTAATCGGATTTTGGATGATTCGAAACAGAGACTTTGAGAAAGCAGAGCA 1569
Db 620 TCGGTAACATTTACTTTCTCTGATCATATACCCGAGCCACCGAGATCATCCCATC 679
QY 1570 GCTCTGAAAAGAGCTGATGATGTCAGAGAAATTTCCACAGCGCATGTCATGACC 1629
Db 680 CAGCTCAACACTCTAGTTGAGAACTGGAACGAGAAAGTACAGAAACCATCATTCAGAGC 739
QY 1630 GAATATGATGAGATACCTCTGACAGGCTTCACTCTATCTCGAGACTGCTTGAAGCGAA 1689
Db 740 GAATACGAGCGATGCGGTGCGGGCTTCACTGATACCTCGGATGTTTACTCGAG 799
QY 1690 GAGTTCCAGTAAATGCTAGACATGATACATGAGTGTGATC-----GCATTGAG 1743
Db 800 GAGTACCAAGAAATTAATCTGACAGAGTACCAAGATGTTGACAGAAAGAAAGCAG 859
QY 1744 TCGATGCGAGCGGAGCATGTTTGAATTTGCCGATTTCCAGACCAATGTTGATCATC 1803
Db 860 TACGTATCGGCGAATCTGAACTTTGACAGACTTATGATGACCCACAGAGGATCATG 919
QY 1804 CGACTAGACGTAACAAGAGGTTTTCACCCGTCGACCGAAAGCCAA 1852
Db 920 CGGCTGTGGGGAACAAAGAGGTGTCCTTCAGAGCAAGGACCCCA 968

RESULT 12

CL478429/c

LOCUS

DEFINITION

SA11.28.D03.v1.SAIL Collection Arabidopsis thaliana genomic clone

ACCENSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 907)

SESSIONS, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D.,

Dietrich, B., Ho, P., Bawaden, J., Ko, C., Clarke, J. D., Cotton, D.,

Bullis, D., Snell, J., Miguel, T., Hutcheson, D., Kimmerly, B.,

Mitrel, T., Karagiri, F., Glazebrook, J., Law, M. and Goff, S. A.

A high-throughput Arabidopsis reverse genetics system

TITLE

JOURNAL Plant Cell 14 (12), 2985-2994 (2002)
 MEDLINE 22356987
 PUBMED 12468722
 COMMENT

Contact: Sessions A
 Applied Trait Genetics
 Syngenta Biotechnology Inc.
 3054 Cornwellis Rd., Research Triangle Park, NC 27709, USA
 Email: allen.sessions@syngenta.com
 ABRIC Stock Number CS801415: T-DNA left border flanking sequences of
 Syngenta Arabidopsis Insertion Library (SAIL) lines are available
 through the Arabidopsis Biological Resource Center (ABRC).
 Sequences represent a pool of amplified genomic regions and not
 single contiguous sequences.
 Class: TDNA tagged.
 Location/Qualifiers

FEATURES
 source
 1..907
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Columbia"
 /db_xref="taxon:3702"
 /clone="SAIL_28.D03.v1"
 /clone_1ib="SAIL Collection"
 /note="T-DNA left border sequences were isolated using a
 modified TAIL-PCR strategy"

ORIGIN

Query Match 8.7%; Score 164.8; DB 9; Length 907;
 Best Local Similarity 59.2%; Pred. No. 5.7e-40;
 Matches 302; Conservative 0; Mismatches 202; Indels 6; Gaps 1;
 Oy 1125 CCGGAGAGATCATGATTCGACATCGAATGGAATTCGTCATCGATGAACACC 1184
 Db CCGGGGGGGGGCGCTTCGCATATATGAACTGCGATCGTATGATGAACCTCG 769
 Oy 1185 TCCCTGTGCTGAC-----ATTGCTTGTATGAGCGCTATCTGAGATGTCGCCA 1238
 Db TCCCTGCGGCTTAACTCTCTTTAGGCAATGCTTGAAGCGGCGCAACAGCGAAGA 709
 Oy 1239 AACATTAGCCAGATGATTAAGATTAACCCAAAGGCCCAAGAGCGCATTCG 1298
 Db ACTGTACAGCGAAGAGGAGTCAACGGGAACTCAGCAAGCGCATTAACAGCGATTA 649
 Oy 1299 TGAGCTCATTCGCCGAGCAAAAACCATGCGAGTGTTCATGTCATGTCATTC 1358
 Db AAGAGCTGATGCGCGGAGCAAAAACCAAGCGCTGATGATGAGATTCGCCAACA 589
 Oy 1359 GCCCGCATCTCATGAAGATGAGCTGCGAATATCTGAGCCACTGACCAATTTGA 1418
 Db ACCGATACCCGTCGCCAAGTGCACGGGAATATTCGCGCCACTGCGGAAACCAAGCG 529
 Oy 1419 TCACTTGTATCCAACTGCGCCCTTTATCATTTGCTAACCTGCGACCGCATATCAG 1478
 Db TAAACTCGACCGACCGCTCCGATCACTGCGCAATGTAATGTTTGGCAGCGCTCAC 469
 Oy 1479 GGATCGGATCTGATCTGTTTGTATGATGATGATTAATCGGATTTTGGATGATTC 1538
 Db CGATACATCAGGATCTCTTTGATGTCTGTCGACCGATTAATTAAGATGATGAT 409
 Oy 1539 TCAACAGAGACCTTGAAGAGCAGAGCGACTTGAAGAGAGCTGATGATGAGCA 1598
 Db CCAAGCGCGGATTTGGTAGCGGCAAGAGATGAGAAAGAACTTCTGCGCTGGCA 349
 Oy 1599 AGAGAAATTCACAGCGCGATGCTCATGAC 1628
 Db GGAGAAATTCGATCAAGCGGATTAATGATGTC 319

RESULT 13
 CD503098 1138 bp mRNA linear EST 12-JUN-2003
 LOCUS CD503098
 DEFINITION CDA60-D07.5' mRNA sequence.
 ACCESSION CD503098

VERSION CD503098.1 GI:31433163
 KEYWORDS EST.
 SOURCE Gasterosteus aculeatus (three spined stickleback)
 ORGANISM Gasterosteus aculeatus

REFERENCE
 AUTHORS Kingsley, D.M., Petchel, C., Balaband, S., Grimwood, J., Dickson, M.,
 Schmutz, J., and Myers, R.M.
 TITLE Expressed sequence tags from Gasterosteus aculeatus
 JOURNAL Unpublished (2003)
 COMMENT Contact: Kingsley, DM
 HMI and Department of Developmental Biology
 Stanford University School of Medicine
 Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
 Tel: 650 725 5954
 Fax: 650 725 7739
 Email: kingsley@cimgm.stanford.edu
 Plate: 60

FEATURES
 source
 1..1138
 /organism="Gasterosteus aculeatus"
 /mol_type="mRNA"
 /strain="Salinas river, CA"
 /db_xref="taxon:69293"
 /clone="CDA60-D07"
 /sex="mixed male and female"
 /tissue_type="heads and internal organs combined"
 /dev_stage="adult"
 /clone_1ib="SHGC-CDA"
 /note="Vector: lambda ZAP Express/PBK-CMV. Site 1: EcoRI
 (5' adaptor); Site 2: XhoI (3' linker primer). The mixed
 organ cDNA library was generated using the ZAP-CMV method
 by Stratagene. First strand cDNA synthesis was primed with
 a 50 bp linker primer containing an oligo dt sequence
 preceded by a synthetic XhoI site. 5 prime adaptors were
 used containing an EcoRI cohesive end. The finished cDNAs
 were inserted in to the ZAP express vector
 unidirectionally in the sense orientation with respect to
 the lacZ promoter of PBK-CMV. An amplified library was
 prepared from approximately 3 million primary clones in
 the lambda ZAP Express vector. In vivo excision was then
 used to generate individual PBK-CMV phagemid clones for
 EST sequencing."

ORIGIN

Query Match 8.3%; Score 158.8; DB 6; Length 1138;
 Best Local Similarity 52.8%; Pred. No. 5e-38;
 Matches 502; Conservative 0; Mismatches 387; Indels 61; Gaps 5;
 Oy 912 CGACACTTCAATTTGGCTACCGGCGTGTCTGATCAAGTTGCCGATCAATTTCTT 971
 Db CGACGATACACTCTACAGAGTGGATCGGACGGTGAAGTTACCAAGACCCAGTTCC 132
 Oy 972 AATAATGGAAGCTTTTCTACTTACCGGTTTGGCAACATGAAGACAGCAGTAGC 1031
 Db CATCAACAAAAGCCCTTCTACTTCCAGAGTAATTAACAGAGACTCTGATATTGG 192
 Oy 1032 TGGCAAGAGCATGACCCAGCATACATGATTCAGATTCACATCATGAATGATTCG 1091
 Db AGCAAAAGGCTGGATGGCCCTCATGGTAGAGACTTTAACTTATTAAGTGGTTGG 252
 Oy 1092 AGCAATTTCTTTCGACTTCACTATATACCGGGAAGAGTGAATGATTCGACAG 1151
 Db GGCAAATCTGTTCCGACACAGCACTTATATGAGAGATCTCTCAAGATGTGTGA 312
 Oy 1152 TGAATGGAATTTGTGATGATGAACACTGCGCTTGGTGTGAACATTCCTTGA 1211
 Db CCGCCATGCGATCTGTGATGAGAGATGTCGCGGCGGTGG----- 354

QY 1212 GGGCGTATCTGAGATGGTGCCCAACAACTTTACGCGAGATGGATTAGATTAAC 1271
 DB 355 -----CATAAAAGACATTGGCAGTTTGGAAAGGCC 385
 QY 1272 CCAGAGGCC--CAACAGCAGCGATTCGTAGCTCATTTGCCGAGACAAAACATATGCC 1329
 DB 386 TCTTAACCCATACCTGCTGTCTATGAGACGACCTGTACGTGGGACAAGAACATATCC 445
 QY 1330 AGGTGTCTCATGTGTCTATTTGCAACGAGCCCGCATTCATTAAGATGAGCTCGCA 1389
 DB 446 TCTGTGTCTCATGTGTCTATTTGCAACGAGCCCGCATTCATTAAGATGAGCTCGCA 505
 QY 1390 TACTTGAGCCATCTGACCAATTGACTCTGCACTTTGATCAACTGCGCTTATTCATTT 1449
 DB 506 TATTTCAAAACCTTGATTAATAACATACCAAGATTTGATTCACACCGGCGCTCATTT 565
 QY 1450 GCTAACGTGCGGACGCGGACATATTCAGCTGATTCGATCTGATCTGTTTATTCAGT 1509
 DB 566 ATCAGACAGC-----TAACTATGCGAGGATTAAGGGGCTCCCTACGTGACGTCATC 619
 QY 1510 TGCATTAATCGGATTTTGGATGTCTTCTCAACAGAGACCTTGAGGAAAGACAGAGCA 1569
 DB 620 TGGTAAACAGTTACTTCTCTGTACATGACCGGCGCACCGGAGTGTATCCCATC 679
 QY 1570 GCTCTGAAAAGAGCTGATGATGAGGAGAAATTCACAGCCGATGTCATGAC 1629
 DB 680 CAGCTCAACATCAGTTTGAAGATGTAAGGAAAGTACAGAAACCCATATCCAGAGC 739
 QY 1630 GAATATGCTGACATACCTTTGACAGCCCTTCACTTATCTCGAGCTGCTTGGAGGAA 1689
 DB 740 GAATACGAGCGGATGCGGTGCGGGCTTCAAGTATCAACCGGTATGTTTACTGAG 799
 QY 1690 GAGTTCGAAGTCAAAATGCTAGACATGACATGAGTGTTC-----GCAATTGAGT 1744
 DB 800 GAGTACCAAGATTAATGCTCTGAGAGCTACCAACAGTTCGACGAAAAGAGAGT 859
 QY 1745 CGATGCGAGCGAGCATGTTTGAACCTTGCCGATTTCCAGAACCACTGGTATCATTC 1804
 DB 860 ACGTATCGGCGAATCATCTGAACTTTGACAGCTTCATGACACCAAGGATTCATGC 919
 QY 1805 GAGTACGAGTAAACAAAGAGGTGTTTCACTCCCTGACCAAGGCCAAG 1854
 DB 920 GCGTGTGGGAGCAAG-AGGTTGTCTTACAGCAGAAAGGACGCAAG 968

 RESULT 14
 CFS21612 740 bp mRNA linear EST 10-SEP-2003
 LOCUS AGENCOURT 15508970 NICH_Ki1 Xenopus laevis cDNA clone
 DEFINITION IMAGE:7012148 5', mRNA sequence.
 CFS21612
 ACCESSION CFS21612.1 GI:34572485
 VERSION
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 Xenopodinae; Xenopus; Xenopus.
 1 (bases 1 to 740)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. Igor Dawid
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing By: The I.M.A.G.E. Consortium (LNLN)
 Clone distribution: NCI-CGAP clone distribution information can be

Found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lnl.gov>
 Plate: LLMJ4721 row: n column: 18
 High quality sequence stop: 670.
 Location/Qualifiers
 1. 740
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:7012148"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Kidney; Vector: pCMV-Sport6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.2 kb. Constructed by Life
 Technologies. Note: This is a Xenopus Gene Collection
 (XGC) library."

ORIGIN

Query Match 8.2%; Score 155.4; DB 7; Length 740;
 Best Local Similarity 53.1%; Pred. No. 5, 1e-37;

Matches 416; Conservative 0; Mismatches 316; Indels 51; Gaps 2;

QY 907 GTAGTCACACCTAACATTTGGCTACGGCGCGCTACTGTCAAGTTGCCGCTACAA 966
 DB 9 GTGAAAGATTTTACGTTTGGCTGTGGAAATTAAGATCTGCACTCTCGAGACCAA 68
 QY 967 TTCTTAATTAATGAAGCCTTTCTACTTTACCGGTTTGGCAACATGAAGACAGCA 1026
 DB 69 TTCTCATCAATGAAGAAACATTTCTACTTCATGCTGTCACAAACATGAGATATAT 128
 QY 1027 GTACGTGCAAGGACATGACCCAGCATATCTGTTCAAGTTTCAACTATGAATAG 1086
 DB 129 GTCAAGGTAAGGACATGACCTGCTACTACTCAATAAGATTTTATCTGTTAAGTGC 188
 QY 1087 ATTGAGCAAAATCTTTTGGACTTCAACTATCTCTTAAGCGGAGGATCATGATTC 1146
 DB 189 CTGTGTCTTACTCTCTTCCGAAACAGCATTTATTCATATGCAAGAAATCATGAGACTT 248
 QY 1147 GCAATGCAATTAATGATTTGCTGATGATGAACACCTGCGTGTGTCGAATATGCC 1206
 DB 249 TGTGTAATTAATGATTTGCTGATGATGAATATGCTGATGATGAATTA----- 301
 QY 1207 TTGATGGCGGTATCTGAGAGTGGGCCCAAAATTTACGCAATGAGATTAACAT 1266
 DB 302 -----ATACCTTAAGTTTGGAAAC 323
 QY 1267 AAAACCAAGAGGCCCAAGCAGCGATTCGTAGCTCATTTGCCGAGACAAAACAT 1326
 DB 324 CAATCTTTAAACCACTTTAATATGATGAGAGAGATGAGTGGCGAGGATTAAGAACGG 383
 QY 1327 GCCAGTGTCTCATGTGTCTATTTGCAACGAGCCCGCATTCATGAATGAGAGTGGC 1386
 DB 384 CTTTCGTTTCTATGTGTGTGTTGCAACGACATCTACCTCCCTGTGCTGGG 443
 QY 1387 GAATCTTTCAGGCACTGACCAATTTGACTGTCATCTTGAATCAACTGCGCCTATTA 1446
 DB 444 TATTTATTTTAAACCGTGAATGATACCTAAGCAACTTGATCCACAGACCGGTGCA 503
 QY 1447 TTGTAAAGTGGGACGCGGACATATACGTGATCGGATCTGATCTGTTGATGTC 1506
 DB 504 T-----AGTTTCTATATCTAATGATTAATGACAGAGTCACTTATGTGATGTG 557
 QY 1507 AGTTGCATTAATCGGATTTTCGATGATTTCTCAACAGAGACCTTGAGAGAGAGAG 1566
 DB 558 ATCTGTGTAATAGTATCTTTCTGTATATATATGATGCGGCACTTGAAATTAACCA 617
 QY 1567 GGAGCTTTGAAAAGAGCTGATGAAGGCAAGAAATTCACAGCGCCGATGTCATG 1626
 DB 618 CTCCAATCAATTAATCAATTTGACAGAGTATGAGAGTATCAAAAGCCTATGATACAG 677
 QY 1627 ACCGAATATGTCAGATACCTTGACAGGCTTCACTCTATCTCGAGATGCTTGAAGC 1686

DB 678 AGTAATATGAGACAGATACAAATTCCTGGGTTTCACAGTGAGCCCCCATGTATGTTCACT 737

QY 1687 GAA 1689

DB 738 GAA 740

RESULT 15

CR442804

LOCUS 689 bp mRNA linear EST 19-JUN-2004

DEFINITION CR442804 XCC-tailbud Xenopus tropicalis cDNA clone TTBa080n17 5', mRNA sequence.

ACCESSION CR442804

VERSION CR442804

KEYWORDS CR442804.1 GI:48968391

SOURCE EST.

ORGANISM Xenopus tropicalis (western clawed frog)

Xenopus tropicalis

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 689)

Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J. Sanger Xenopus tropicalis EST project 2001 (2004)

Unpublished (2004)

Contact: Croning MDR

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS SEQUENCE ID: TTBa080n17.plksp6

This sequence is from a Xenopus Gene Collection (XGC) library constructed by Nigel Garrett.

Seq primer: Sp6.

FEATURES

source

Location/Qualifiers

1..689

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/db_xref="taxon:8364"

/clone="TTBa080n17"

/dev_stage="tailbud (stage 28-30)"

/lab_host="Escherichia coli DH10B."

/clone_lib="XGC-tailbud"

/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from Sug of poly A+ RNA from tailbud. EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 7.9%; Score 150.6; DB 7; Length 689;

Best Local Similarity 53.9%; Pred. No. 1.6e-35;

Matches 392; Conservative 0; Mismatches 284; Indels 51; Gaps 2;

QY 950 AGGTTGCGGGTCAATCTTATTAATGAAGCCTTTCTAATTACCGGTTTGGCA 1009

DB 4 AGCTCTCGAAGCAATTCCTCATGAGAAACCATTTCTTCAGTGTCAACA 63

QY 1010 AACATGAAGACACAGCACTACGTCGAAAGAGACATGACCATACATGCTTCAGATT 1069

DB 64 AACATGAGACTATGATGTCAGAGAAAGACTAGACTGCTCACTAATTTGTAAGAGATT 123

QY 1070 TCCAACTCATGAATAATGATGAGCAAAATCTTTGSACTTCACTATCTTACGCGG 1129

DB 124 TTTATCTGTGAAGTGGCTTGGCTGCTACTCTTCGACACAGCCATTTATCTTATGAG 183

QY 1130 AAGAGTCAATGATTTTCAGATGAAATGGAATGTCGTGATGATGAACAACCTGCGG 1189

DB 184 AAGAAATCAATGACCTTTGTGATTAATATGCAATTTGGATTTGATGAATGCCGTGAG 243

QY 1190 TTGGTCTGAACATTCCTGATGAGGCGTATCTGAGAGTGTCGCCCAACAACATTTAAGC 1249

DB 244 TCGGTATTA-----ATACC 258

QY 1250 CAGATGATTAAGATTAACCAAGAGGCCCAAGACGAGGATTCGTGAGCTATTG 1309

DB 259 CCGAAGTTTGGAAACCAATCTTTAAACACCATTTAATATGATGAGAGTTGTGC 318

QY 1310 CCGAGACAAAACCATGCGCACTGTTGTATGATGCTATGTCACAGCCGCACTGC 1369

DB 319 GCAGGATTAACCAACCGGCTTCGTTGTATGATGCTATGTCACAGCCGCACTGC 378

QY 1370 ATGAAGATGAGTGCAGAAATCTTGAGCCACTGACCAATTTGACTGCTCACTGATC 1429

DB 379 AGCTCCCTGTGCTGGGTATTTATTTAAACGGTATGTTGATGATGATGATGATG 438

QY 1430 CAATGCGCCCTATTAATTTGCTTAACGTGCGCAAGGCGACATATACGCTGATGATCT 1489

DB 439 CACCCGACAGAGACAT-----ATGTTCCAAATCTAATATGAACATGACCAAGTGC 492

QY 1490 CTGATCTGTTGATGCTCACTGATGATTAATCGGATTTTGGATGATTTCTCAACGAG 1549

DB 493 CACCTATGATGATGATGATTTGTTAAACAGTTACTTTCTGTTATCATGATGATGCGG 552

QY 1550 ACCTGAGAGAGAGAGGAGGAGCTTTGAAGAAGAGCTGCATGATGAGCAAGAAATTC 1609

DB 553 ACCTGAAATTTATCCAACTCAACTCAATGATGATGATGATGATGATGATGATGATG 612

QY 1610 ACAGGCGCATGCTCATGACCGAATATGTCAGATACCTTGAGGCTTCACTATTC 1669

DB 613 AAAAGCTATGATACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 672

QY 1670 TCGGACT 1676

DB 673 CCCCAT 679

Search completed: March 22, 2005, 18:00:12

Job time : 5670 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: March 22, 2005, 18:23:41 ; Search time 4458 seconds

(without alignments)
5413.357 Million cell updates/sec

Title: US-10-757-093-4

Perfect score: 3354
Sequence: 1 MFLNGLSLSLAPSLGTP.....RKPKAAHSLRWRWSIDKN 634

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=alp
-Q/cg2n2.1/USPTO.spool.p/US10757093/runat_18032005_164456_27770/app.query.faeta_1.775
-DB=EST -QFMT=fastap -SUFPIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0
-UNIT=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10757093.CGCN_1_1.5180.0/runat_18032005_164456_27770 -NCPU=6 -ICPU=3
-NO_MMWP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:.*
1: gb_est1:.*
2: gb_est2:.*
3: gb_hic:.*
4: gb_est3:.*
5: gb_est4:.*
6: gb_est5:.*
7: gb_est6:.*
8: gb_g881:.*
9: gb_g882:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1267	37.8	2473	3	AY321342 Rattus no
2	1253.5	37.4	2095	6	CD014094
3	1175	35.0	1995	6	CD014092
4	1168.5	34.8	2274	3	AK041058
5	1045.5	31.2	1853	6	CD014093
6	946	28.2	1377	3	CR593823
7	803	23.9	1124	6	CD503076
8	797.5	23.8	1051	5	BX363460
9	797	23.8	921	9	CL486845

10	770.5	23.0	1138	6	CD503098	CD503098 CDA60-D07
11	726.5	21.7	857	5	BX745933	BX745933 BX745933
12	712.5	21.2	811	9	CL679241	CL679241 PR10125C
13	705.5	21.0	906	5	BO941196	BO941196 AGENCOURT
14	689	20.5	1055	4	BM557676	BM557676 AGENCOURT
15	683	20.4	877	7	CF406520	CF406520 CH3#042_G
16	681.5	20.3	767	7	CF255373	CF255373 mdv1127_c
17	674.5	20.1	740	7	CF521612	CF521612 AGENCOURT
18	671.5	20.0	816	7	CK467358	CK467358 938649 MA
19	671.5	20.0	878	7	CK155220	CK155220 942501 MA
20	661.5	19.7	914	6	CB203472	CB203472 AGENCOURT
21	656	19.6	1063	5	BX401772	BX401772 BX401772
22	649.5	19.4	915	5	BO678153	BO678153 AGENCOURT
23	648.5	19.3	846	6	CA453907	CA453907 AGENCOURT
24	647.5	19.3	802	5	BU357212	BU357212 603474094
25	643	19.2	733	7	CN254047	CN254047 170006000
26	637	19.0	736	5	BP454056	BP454056 BP454056
27	633.5	18.9	791	5	BU239978	BU239978 603323761
28	630.5	18.8	689	7	CR442804	CR442804 CR442804
29	622.5	18.6	725	5	BU315788	BU315788 603851444
30	621	18.5	566	7	CV169266	CV169266 rsmex1_01
31	620.5	18.5	900	6	CA489317	CA489317 AGENCOURT
32	609.5	18.2	799	5	BU36987	BU36987 603585124
33	609	18.2	731	4	BG121498	BG121498 602352830
34	605.5	18.1	731	7	CV105482	CV105482 AGENCOURT
35	601.5	17.9	790	7	CO559438	CO559438 AGENCOURT
36	601	17.9	680	7	CK836522	CK836522 4061270 B
37	591	17.6	898	5	BU179563	BU179563 AGENCOURT
38	590	17.6	962	5	BQ684422	BQ684422 AGENCOURT
39	587.5	17.5	773	7	CF748482	CF748482 UI-M-HJ0-
40	585.5	17.5	757	4	B1851705	B1851705 603377940
41	583.5	17.4	847	5	BX621253	BX621253 BX621253
42	582	17.4	925	5	BO679162	BO679162 AGENCOURT
43	578.5	17.2	624	4	BI445636	BI445636 deeb305
44	577.5	17.2	673	4	CB841568	CB841568 M5E-2121
45	573	17.1	857	6	CB559443	CB559443 AGENCOURT

ALIGNMENTS

RESULT 1
AY321342
LOCUS AY321342 2473 bp mRNA linear HTC 16-JUN-2003
DEFINITION Rattus norvegicus Ac2-223 mRNA, complete cds.
ACCESSION AY321342
VERSION AY321342.1 GI:32527744
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
HTC.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 2473)
XU,C.S., LI,W.Q., LI,Y.C., WANG,L., WANG,S.F., HAN,H.P., WANG,G.P.,
CHAI,L.Q., YUAN,J.Y., YANG,K.J., YAN,H.M., CHANG,C.F., ZHAO,L.F.,
MA,H., SHI,J.B., RAHMAN,S., WANG,Q.N. and ZHANG,J.B.
Liver regeneration after PH
Unpublished
2 (bases 1 to 2473)
XU,C.S., LI,W.Q., LI,Y.C., WANG,L., WANG,S.F., HAN,H.P., WANG,G.P.,
CHAI,L.Q., YUAN,J.Y., YANG,K.J., YAN,H.M., CHANG,C.F., ZHAO,L.F.,
MA,H., SHI,J.B., RAHMAN,S., WANG,Q.N. and ZHANG,J.B.
Direct Submission
Submitted (11-JUN-2003) Henan Bioengineering Key Lab, Henan Normal
University, NO. 148 Jianshe Road, Xinxiang City, Henan 453002, P.R.
China

FEATURES

source
1..2473
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
41..2467
/note="liver regeneration related protein LRRC134"

```

/codon_start=1
/product="Ac2-223"
/protein_id="AAP86274.1"
/db_xref="GI:32527745"
/translation="MKEKDFKQEGVEKYLMDLNSVYLQANERDSREAKRPSRWTAI
PMSRSRRCWFWVLGQLCSALAQGMLEPKETPRELKVLDGLMSPADYNNR
LQGFKEQWYRQPLRSRSGPTLMDPVPSSEFNDITQEALEPLIGMWYBEALPRLMIO
DIDRVLVINSAAHYAVVWVNGIHVHEGGHLPFADITIKVNGSGITPTPTIAL
NNTLPEYLPPTIYKTDPSMYPKGYFVODISPDPRYAGLHNSVLYTPTTID
LVTVDVDRDGLVNWVMSVQSGDHPQLEFVRLIDEDGKITVARGNBSGLVPRAHIM
WPKMHEHRAVYXSLEVTMTTPESVDTPTLPVGRITVAIVKSKFLNGKPYQGVN
KHEEDIRKRGFDMPLIKDFNLKMLDANSFRSHYSEVLQCDRYGIVVIDEC
PEVGI VLPQSGFNVSRLRHLLEVMDLVANDKHPAVVMSVANSPEVSLKAGYFFK
LVQGLKTSIDSPRTGVALVNVHAYVAVIVICVNSLSWYHGLVYLOLOTFEN
WYKMYOKPIIOSEYGDAVSGLEDPRPMSSEYOTALLENVYHILDEKREYVIGEL
IWNFADFMNOSPLRVTKNGKGIETRONRPMAPAIERWRRIANETRGILQVNSF
VYDDEKEVOHICGDADAKILQHKYQVNGODSVIAGALICSTCTNRHAELOCKST
VPSDAGAOERQQLSIOATGMATVITINACSHFQVRDLKTHVANSIAVFKDPRLS
CLIDISAGSA"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 1.66e-130 Length: 2473
Score: 1267.00 Matches: 266
Percent Similarity: 57.74% Conservaive: 96
Best Local Similarity: 42.42% Mismatches: 193
Query Match: 37.78% Indels: 72
DB: 3 Gaps: 13

```

US-10-757-093-4 (1-634) x AV321342 (1-2473)

```

QY 42 ProGlnaGThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu 61
DB 269 CCGAAGGAGACCCCTTCGGGAACTCAAGGTCGTGACGCGCTGTGACCTTCCTG--- 325
QY 62 AlaSerGlyLeuAsnAspThrAlaGlnPro-----TrpThrAlaProLeu 76
DB 326 GCCGATTACTGACCAATCGCTTCACAGCGCTTCGAGAGACAGTGTGATCCGGACCGCTA 385
QY 77 ProLysGly-----LeuGluCysProValProAlaSerTrpAsnAspIlePheAla 93
DB 386 CGGAGAGTCGGGCCCAACCTTGACATGCCGTCCTTCAGCTTCATATGACATCACCAA 445
QY 94 SerArgGluLeuHisAspHisValGlyTrpValTrpTrpGlnArgGluValIleValPro 113
DB 446 GAAGCAGACGCTTCGACATTCATGCGTGGGTGGTGGATGAAACGGGAAACAGGCTTCCA 505
QY 114 LysGlyTrpSerGlnGlu-----ArgTrpLeuValArgAlaGluSerAlaThrHis 130
DB 506 CAGCGATGAGACCCAGGACCCGACAGAGAGGTGTGTGATGAATCAACGCCCATTTAC 565
QY 131 HisGlyValArgIleTrpValAsnAspArgLeuValAlaGluHisValGlyLeuTrpPro 150
DB 566 TATGCACTGTGTGGGTGAATGGGATTCATGTGTGAAACATGAGGAGGACGACCTCCCC 625
QY 151 PheGluAlaAspValThrGluLeuValAlaProGly-----GluLysPheArgLeuThr 168
DB 626 TTGAGAGCTGACATCACCAAGGTGTCCAGAGTGGGCCCTTACCACTTCGGGGTACC 685
QY 169 IleGlyValAsnAsnGluLeuThrHisGluTrpIleProProGlyLysIleThr----- 186
DB 686 ATGCGCATCAACAACACCTGACCCCTTATACCTTCACCGGGGACCATTTGTACAAAG 745
QY 187 -----ThcGlyAsnAlaThrGlyLysArgIleGlnThrTrpGlnHisAspPheTrp 203
DB 746 ACTGATCTTCATGATCCCAAGGTTACTTCGTCCAGAGACATAGCTTGACATTTCTTC 805
QY 204 AsnTrpAlaGlyLeuAlaArgSerIleTrpLeuTrpSerValProGlnGlnHisIleGln 223
DB 806 AACTATGCGGGGTGACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 865
QY 224 AspIleThrValValThrAspValAspGlyAspAsnGlyLeuIleAsnTrpGluValGlu 243

```

```

DB 866 GATATCACTGTGACCACTGACGTGACCCGGGACGTTGGGTGTGAACTAGATTCT 925
QY 244 ValAlaAsnGlnThrGlnGlnIleGlnIleSerValIleAspGluAspGlyAlaIle 263
DB 926 GTCCAGGGCAGTACCATTTCCAGCTTAAAGTGGCTTCTGTGATGAGATGGCAAAAT 985
QY 264 ValAlaLeuAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTrpGln 283
DB 986 GTGGCCCGTGAACACAGGAATGAGGCTCAATTAGGTGCCCGGTCCACCTCTGTGG 1045
QY 284 Pro-----GlyAlaAlaTrpLeuTrpGlnLeuGlnValAsnIleValGly 298
DB 1046 CTTTACCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1102
QY 299 SerSerGlyAspValValAspThrTrpAsnLeuAlaThrGlyValArgThrValLysVal 318
DB 1103 ACACGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1162
QY 319 AlaGlySerGlnPheLeuIleAsnGlyLysProPheTrpPheTrpGlyPheGlyLysHis 338
DB 1163 ACAAGAGCAAGTTCCTCATTAATGGAAGCCTTCTCATTTCCAAAGCGGTCAACAGCAT 1222
QY 339 GluAspThrAlaValArgGlyLysGlyHisAspProAlaTrpMetValHisAspPheGln 358
DB 1223 GAGGATTCAGATATCCAGAGGAGAGGCTTCACATGAGCTGTGTGATTAAGATTTCAAC 1282
QY 359 LeuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisArgTrpProTrpAlaGluGlu 378
DB 1283 CTCCTCCGTGTGCTCGGGGCAAAATCTTCTGTACACACCATATCCCTATCCGGAAGAG 1342
QY 379 ValMetAspPheAlaAspArgAsnGlyIleValValIleAspGluTrpProAlaValGly 398
DB 1343 GTACTTCAGCTGTGTGACCCGATGGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1402
QY 399 LeuAsnIleAlaLeuMetGlyValSerGluSerGlyAlaProGlnThrPheThrProAsp 418
DB 1403 ATCGTGTG-----CCCGAGATTTT----- 1423
QY 419 AlaIleAsnAspLysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArg 438
DB 1424 -----GCCAAGTGTCTCTTGGCACCCACTTAAGGTGTGTGACGAGCTGTGGCAGG 1477
QY 439 AspTrpAsnHisAlaSerValValMetTrpSerIleAlaAsnGluProAlaSerHisGlu 458
DB 1478 GACAAATAATCACCTGGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1537
QY 459 AspGlyAlaAspGlyTrpPheGlu----- 466
DB 1538 AAACCTCGGATTTACTTCAAGCTTATGTCTTCCAAAGCCTGAAGACATCAGATTCC 1597
QY 467 ProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnVal 486
DB 1598 CCAAGAACTGAGTTACAGCTCTTGAACCTTACCAAGCCCGGTACGTG----- 1645
QY 487 GlyThrAlaThrTrpGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsn 506
DB 1646 -----GACGTGATTTGTGTGAAC 1663
QY 507 ArgTrpPheGlyTrpTrpSerGlnThrGlyAspLeuGlnGluAlaGluAlaLeuGlu 526
DB 1664 AGTTACTTATCCGTGGTATCTAGCTACGGGACATCTGGAGGTATTCATGTCGAGCTGACT 1723
QY 527 LysGluLeuHisGlyTrpGlnGluLysPheHisArgProIleValMetThrGlnTrpGly 546
DB 1724 AGCCAGTTTGAGAACTGTGATTAAGATGTCAAGAAAGCAAATTATACAGAGCAAGTGTGA 1783
QY 547 AlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGlnGluPheGln 566
DB 1784 GCAAGAGCGGTCTCGGGGTCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1843
QY 567 ValGlnLeuAspMetTrpHisArgValPheAsp-----ArgIleGlnSerMetAla 584
DB 1844 ACAAGCTCTCTGGAAGATTAATCATTTGATCTGTGATGAGAAACGAAGAAATATGTCTATC 1903

```

QY 585 GUGUUNhValTTPhnphehAaApheGlnThrAsnLeuGlyLeileArgValaAp 604
DB 1904 GAGAGCTACTCTGGAAATTTTGGCTGACTTATGACAAACAGTCCAGCAGGAGTAAACA 1963
QY 605 GUAAnUyLeuAGlyValPheThrArgAspArgLyPProLyAlaAlaAnhIserLeu 624
DB 1964 GGAACAAAGAGGGATCTTCACTGACAGAAACCCCAAGATGAGCCTTCATTG 2023
QY 625 ArgAlaArgTTPhrSerIle 631
DB 2024 CGAGAGAGATCTGAGAGATT 2044

RESULT 2
LOCUS CD014094 2095 bp mRNA linear EST 21-OCT-2003
DEFINITION 90135027 Single gene library Homo sapiens cDNA, mRNA sequence.
ACCESSION CD014094
VERSION CD014094.1 GI:37777623
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2095)
Jin, P., Fu, G. K., Wilson, A. D., Yang, J., Chien, D., Hawkins, P. R.,
Au-Young, J. and Stuve, L. L.
PCR isolation and cloning of novel splice variant mRNAs from known
drug target genes
Genomics 83 (4), 566-571. (2004)
CONTACT: Jin, P.
Incyte Corporation
3160 Porter Drive, Palo Alto, CA 94304, USA
Tel: 650 621 8639
Fax: 650 621 8965
Email: pj@incyte.com.
Location/Qualifiers
1..2095
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="Single gene library"
/note="Vector: pDrive Cloning Vector; RT-PCR was performed
using gene-specific primers flanking the open-reading
frame. PCR products were subcloned into pDrive Cloning
Vector and sequenced completely using M13 forward and
reverse primers. Sequencing gaps were closed by
re-sequencing using primers flanking the gapped areas."

ORIGIN

Alignment Scores:
Pred. No.: 4.21e-129 Length: 2095
Score: 1253.50 Matches: 266
Percent Similarity: 57.51% Conservative: 90
Best Local Similarity: 42.97% Mismatches: 187
Query Match: 37.37% Indels: 77
DB: Gaps: 13

US-10-757-093-4 (1-634) x CD014094 (1-2095)

QY 42 ProGlnArgThrSerSerArgGluValAsnLeuAspGlyLeuTrpLysPheAlaLeu 61
DB 101 CCCAGAGAGGCCGCTGCGGAGTGCAAGAGCTGACCGCCTCTGAGACTTC----- 154
QY 62 AlaSerGlyLeuAnAspThr-----AlaGlnProTrpThrAlaPro 75
DB 155 CGGCGCGACTTCTGTGCAACGACGCGGGCTTCGAGAGCAGTGTGTCGCGCGCG 214
QY 76 LeuProLysGly-----LeuGluCysProValProAlaSerTyTrpAnAspIlePhe 92
DB 215 CTGTGGAGATGACGCCACCGGACATGCGAGTTCCCTCCAGCTTCATATGACATCAGC 274
QY 93 IleSerArgGluIleHisAspHisValGlyTrpValTyTrtyGlnArgGluValIleVal 112

DB 275 CAGAGCTGGCGTCTGGCGGCAATTTTGTGGCTGGTGTGTAAGAACGGAGGATCTG 334
QY 113 ProLyAGlyTrpSerGlnU-----ArgTyLeuValArgAlaGluSerAlaThr 129
DB 335 CGGAGCGATGAGCCAGGACCTGCGCACAAAGTGTCTGTGAGTGTGGCAATGCCAT 394
QY 130 HbHAGlyArgIleTyValAsnAspArgLeuValAlaGlnHbValGlyGlyTyTrh 149
DB 395 TCCTATGCCATGTGTGGTGAATGGGTGACAGCCTGAGACATAGAGGGGGCTACCTC 454
QY 150 ProPheGluAlaAspValThrGluLeuValAlaProGly-----GluTyPheArg 166
DB 455 CCTTGAGGCGGAGCATGACAGAACTGTGTCAGGTGGGCCCCCTGCCCTCCGCGTCCGA 514
QY 167 LeuThrIleGlyValAsnAnGluLeuThrHisGlyThrIleProProGlyLysIle--- 185
DB 515 ATCACTATGCGCATCAACAACACATCACCCCAACCCCTGCGCACAGGAGCATCCAA 574
QY 186 -----ThrThrGlyAsnAlaThrGlyLysArgIleGlnThrTyTrhHisAsp 201
DB 575 TACCTGACTGACACTTCCAAAGTATCCCAAGGTACTTGTCCAAACATATTTTGAC 634
QY 202 PheTyAsnTrpAlaGlyLeuAlaArgSerIleTrpLeuTySerValProGlnGlnHis 221
DB 635 TTTTCACTACGCTGAGTGGAGCGGTCTGTATCTGTATGACAGAACCCACCACTAC 694
QY 222 IleglnAspIleThrValValThrAspValAspGlyAspAsnGlyLeuIleAsnTyTrh 241
DB 695 ATCGATGATCATCCCTGACCAACAGCGTGAGCAAGAGATGGCGGTGGATTCAG 754
QY 242 ValGluValAlaAsnGlnThrThrglyGlnIleglnIleSerValIleAspGluAspGly 261
DB 755 ATCTGTGTAAGGCGAGTAACTGTTCAAGTTGAAGTGCCTTTGTGATGACAGAAAC 814
QY 262 AlaIleValAlaLysAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeu 281
DB 815 AAAGTGTGGCGAATGGACTGGAGACCGAGGCAACTTAAGGTGCGAGGTGCACCTC 874
QY 282 TrpGlnPro-----GlyAlaAlaTyLeuTyTrhGlnGluValAsnIle 296
DB 875 TGGTGCCGTACTGTCGACGACCAACCCCTGCTATCTGTATTCATGAGGAGCGAGCTG 934
QY 297 ValGly-----SerSerGlyAspValValAspThrTyAsnLeuAlaThrGlyValArg 314
DB 935 ACTGACAGACGTCATGCGGCTGTGTGATCTTACACATCCCTGTGGGAGATCCG 994
QY 315 ThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysAspProPheTyTrhTrhGly 334
DB 995 ACTGTGCTGTACCAAGAGCCAGTTCATCAATGAGAAACCTTCTATTTCCACGGT 1054
QY 335 PheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyTrhMetVal 354
DB 1055 GTCAACAAACATGAGATCGGACATCCAGGAGAAAGGCTTCGACGCGCGCTGCTG 1114
QY 355 HisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisTyTrp 374
DB 1115 AAGGACTTCAACTGCTGTGCTGTGGTGGTGGCCAGCTTTCGTCACAGCCATCACCC 1174
QY 375 TYTAlaGluGluValMetAspPheAlaAspArgAsnGlyTleValAlaIleAspGluThr 394
DB 1175 TATGACAGAGAGTATGACATGTGTGACCGGTATGAGATGTGTATTCATGATGATG 1234
QY 395 ProAlaValGlyLeuAnIleAlaLeuMetGlyValSerGluSerGlyAlaProGlnThr 414
DB 1235 CCGCGCGTGGCGCTGCGCTG-----CCGAGTTTC 1264
QY 415 PheThrProAspAlaIleAsnAspLysTrhGlnGluAlaHisGlyGlnAlaIleAspGlu 434
DB 1265 TTC-----AACACGTTTCTGTGATCACCACTGACGATGAGAAAG 1309
QY 435 LeuIleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAlaAsnGluPro 454

QY	Db	1310	GTGGTGGGTGGGGAACAAGAACCAACCCCGCGCTGTCATGTGGTCTGTGGCAACAGACGCT	1369
QY	455	AlaserhisgcluarpsglvalargvgutyrphgluProleuThrasnleuThrArgGln	474	
Db	1370	GGCGCCACCACTAGAAATCTGCTGGCTACTT--GAA-----	1404	
QY	475	LeuaspProthraArgProileThrPheAlaasnValGlyThrAlaThrGlyGlnLeuasp	494	
Db	1405	-----GCTCCCGTAT-----	1413	
QY	495	ArgIleSerAspLeuPheAspValSerCysIleasnArgTyrPheGlyTyrTrpSerGln	514	
Db	1414	-----GTGGATGTGATCTGTTGAACACTACTACTCTTGGATACAGAC	1458	
QY	515	ThrgIAspRleglgluAlaGlyAlaAlaLeuGlnGlyIleuHisGlyTyrGlnGlu	534	
Db	1459	TACGGCACCTGGAGAGTTGATTCAGCTGGACGCTGGCCACCCAGTTTGAGAACTGGTATAG	1518	
QY	535	LyvPheHisArgProIleValMetThrgIuTyrGlyAlaAspThrIleuAlaGlyLeuHis	554	
Db	1519	AAGTATCAGAAACCCCTATTTCAGAGCAGATGTAGCAGAAACGATTCGAGGGGTTTCAC	1578	
QY	555	SerIleLeuGlyLeuProTrrpSerGlnGluPheGlnValGlnMetLeuAspMetTyrHis	574	
Db	1579	CAGATTCACCTGTGATGTTCATCTGAAAGATACAGAAAGAGTCTGTAGACAGATACCAT	1638	
QY	575	ArgValPheAsp-----ArgIleGluSerMetAlaGlyGlnHisValTrrpAsnPheAla	592	
Db	1639	CTGGGTCGTGANTCAGAAACCGCAGAAATATAGCTGTGGAGAGCTCATTTGGAATTTTGCC	1698	
QY	593	AspPheGlnThrAsnLeuGlyIleIleArgValAspIlyAsnIlyGlyValPheThr	612	
Db	1699	GATTTTCATAGCTGGAACACTGCACGACGAGAGTGCTGGGGAATTAAGGGGANTTCACT	1758	
QY	613	ArgAspArgIlyProIlyAsnAlaAlaHisSerIleuAlaGlyAlaArgTrrpThrSerIle	631	
Db	1759	CGGCAAGACACACCAAAAAGTGCAGCGCTTCCTTTTGGAGAGAGATACGGAAGATT	1815	
RESUIT 3				
LOCUS	CD014092	1995 bp	mRNA	linear
DEFINITION	901334967	Single gene library	Homo sapiens	cDNA, mRNA sequence.
ACCESSION	CD014092			
VERSION	CD014092.1	GI:37777621		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Emukayrola; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
TITLE	1 (bases 1 to 1995)			
JOURNAL	Jin,P., Fu,G.K., Wilson,A.D., Yang,J., Chien,D., Hawkins,P.R.,			
COMMENT	PCR isolation and cloning of novel splice variant mRNAs from known			
	Genomics 83 (4), 566-571 (2004)			
	Contact: Jin, P.			

FEATURES
SOURCE

"/clone_id="Single gene library"
/note="Vector: pRivite Cloning Vector; RT-PCR was performed using gene-specific primers flanking the open-reading frame. PCR products were subcloned into pRivite Cloning Vector and sequenced completely using M13 forward and reverse primers. Sequencing gaps were closed by re-sequencing using primers flanking the gapped areas."

Alignment Scores:	2,62e-120	Length:	1995
Pred. No.:	1175.00	Matches:	251
Score:	55.92%	Conservative:	89
Percent Similarity:	41.28%	Mismatches:	181
Best Local Similarity:	35.03%	Indels:	89
Query Match:		Gaps:	12
DB:			
US-10-757-093-4 (1-634) x CD014092 (1-1995)			
QY	42 ProGlnArgThrSerSerArgGluLeuValAlaSerLeuAspGlyLeuTrpLysPheAlaLeu 61		
DB	101 CCCGAGGAGAGACCCTCGCGCGGAGTGCAGAGAGCTGACGCGCTCTGAGACTTC----- 15		
QY	62 AlaSerGlyLeuAsnAspThr-----AlaGlnProTrpThrAlaPro 75		
DB	155 CGCGCCGCACTTCTGTGACACCGACGCGGCGCTTCGAGGAGCAGTGCTGACCGGCGCG 214		
QY	76 LeuProLysGly-----LeuGluCyPProAlaProAlaSerTrpAsnAspIlePhe 92		
DB	215 CTGTGGAGATCAGGCCCCACCGTGAATGCAATGCACTTCCCACTTCAATGATCATGACAGC 274		
QY	93 ILeSerArgGluIleHisAspHisValGlyTrpValTrpTrpGlnArgGluValIleVal 112		
DB	275 CAGGACTGGCGTCTGCGGAGCATTTTGTGCGTGGGTGTGATACACCGGAGAGTATCTCG 334		
QY	113 ProLysGlyTrpSerGlnGluArgTrpLeuValArgAlaGluSerAlaThrHisGly 132		
DB	335 CCGGAGCGATGACCCAGGAC----- 355		
QY	133 ArgIleTrpValAsnAsnArgLeuValAlaGluHisValGlyGlyTrpTrpProPheGlu 152		
DB	356 -----CTGCGCACAGAGTGTG----- 373		
QY	153 AlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsn 172		
DB	374 -----CTGAGGATTTGGCAATGGCC 391		
QY	173 AsnGluLeuThrHisGlnThrIleProProGlyLysIleThrThrGlyAsnAlaThrGly 192		
DB	392 CATTC-CTATGCATGCTGTATCCAG-GGTACTTT----- 428		
QY	193 LysArgIleGlnThrTrpGlnHisAspPheTrpAsnTrpAlaGlyLeuAlaArgSerIle 212		
DB	429 -----GTCCAGAACACATATTTGACTTTTCAACTACGCTGAGCTGAGCGGTCTGTA 482		
QY	213 TrpLeuTrpSerValProGlnGlnHisIleGlnAspIleThrValValThrAspValAsp 232		
DB	483 CTTCTGTACAGACACCCACCACTTACATGATGATCATCACCGTACACCAACGACTGGAG 542		
QY	233 GlyAspAsnGlyLeuIleAsnTrpGlyValGluValAlaAsnGlnThrThrGlyGlnIle 252		
DB	543 CAAGGACAGGGCTGTGAAATTACACAGATCTCTGTCAAGGGCAGTAACCTGTTCAAGTTG 602		
QY	253 GlnIleSerValIleAspGluAspGlyAlaIleValAlaLysPheValGlnGly 272		
DB	603 GAAAGCGCTCTTTGGATGCAAAACAAAGTCTGTGGATGGGATGGGAGCTGGACCAAGGCG 662		
QY	273 ThrValThrIleProSerValIleLeuTrpGlnPro-----GlyAlaAla 287		
DB	663 CAACCTTAATGGCCAGAGGTCAAGCTCTGTGTGGCGCCGATACCTGATGACAAAGCCCTGCGC 722		
QY	288 TyrLeuTrpGlnGluGlnValAsnIleValGly-----SerSerGlyAspValValAsp 305		
DB	723 TATCTGATTCATTGAGGTGACGACTGATGCAACAGACGTCACTGGGGCGCTGTCTGAC 782		
QY	306 ThrTrpAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIle 325		
DB	783 TTTCACACACTCCCTGTGGGAGATCCGACATGTGGCTGTACCAAGACCAAGTTCCTATC 842		
QY	326 AsnGlyLysProPheTrpThrGlyPheGlyLysHisGluAspThrAlaValArgGly 345		

Db	843	AAATGGAAACCTTTCTATTTCACACGGGTGCAACAGATGAGATGGCGACATCCGAGGG	902	ORGANISM	Mus musculus
Qy	346	LYSGLYHIAAPRPROALATYRMECVALLHIAAPRHEGLMEUETLSTPILGLYALA	365	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Db	903	AAAGGGCTTCACATCGCCGCTGCTGGTGAAGAGACTTCAACCTGCTTGGCTGGTGGCC	962	AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Qy	366	AAPSRPHEARGTHRSERHISTATYRPROYALAGLUGLVALMECAPRHEALAAPARG	385	TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Db	963	AACGCTTCCGTCACACCCCTACCTTATGCAAGAGATGATGCAAGATGTGACCCG	1022	JOURNAL	1 Carninci, P. and Hayashizaki, Y.
Qy	386	ASNGIYILEVALIILEAPGLUTHRPROALAVALGLEUASNLLEALALEUMETGLY	405	MEDLINE	High-efficiency full-length cDNA cloning
Db	1023	TATGGATGTGTGCTCATCGATGATGTCGCCGGGTGGCGCTGGCGCTG	1070	PUBMED	Meth. Enzymol. 303, 19-44 (1999)
Qy	406	VALSERGLUSERGLYALAPROGLINTHRPHEATHPROAPALALEASAPLYSTHRIN	425	REFERENCE	99279253
Db	1071	-----CCGAGATCTTC-----AACCAAGTTCTCTG	1097	AUTHORS	10349636
Qy	426	GLUALHIELYAGINLAILEARGGLULEUILEALARGAPLYSAPENHIAIASERVAL	445	TITLE	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Db	1098	CATCACCACTGACGATGAGAGAGATGTGCTGCTGAGCAAGACACCCCGCGT	1157	JOURNAL	Itch, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Qy	446	VALMETTPSERILEALANGIUPROALASERHISGLIUALAPGLYALARGIUTYRPH	465	MEDLINE	Normalization and subtraction of cap-trapper-selected cDNAs to
Db	1158	GTGATGTGCTGTGGCGCAACGAGCTGCTCCCACTAGATCTGTGCTACTCTTG	1217	PUBMED	prepare full-length cDNA libraries for rapid discovery of new genes
Qy	466	GLUPROLEUTHRAENLEUTHRARGINLEUAPPROTHARGPROILETHRPHLEALAE	485	REFERENCE	Genome Res. 10 (10), 1617-1630 (2000)
Db	1218	AAGATGTGATCCCTGACACCAATCTTGTGACCCCTCCGCGCTGTGACCTTT	1271	AUTHORS	20499374
Qy	486	VALGILYTHRALATHRYRGINLEUAPARGILSERASPLEUHEAPVALSERVAL	505	TITLE	11042159
Db	1272	GTGAGCACTCTACTACTGACGACAGACAGAGGAGCTCCGATGTGATGTGATCTTTG	1331	JOURNAL	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Qy	506	ASNARGTYRPHGLYTRPYRSEGINTHRGLYASPLEUGLUGLUALAGLUALALEU	525	REFERENCE	Komno, H., Akiyama, J., Nishi, K., Kitsuai, T., Taahito, H., Itoh, M.,
Db	1332	AACAGTACTACTCTGTATGACGATGAGGAGCCTGAGTGTGATGATGATGATGATG	1391	AUTHORS	Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Qy	526	GLULYRGLULEUHIAGLYTRPGINGLULYRPHHIAHRPROILEVALMECTRGLYR	545	TITLE	Yamamoto, K., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Db	1392	GCCACCACCTTGAGAACTGTATTAAGAGTATACAGAACCCATTAATTCAGACGAGTAT	1451	JOURNAL	Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Qy	546	GLYALAPRTHREULAGLYLEUHIASERILELEUGLYLEUPROTTPSERGLUGLUPHE	565	MEDLINE	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,
Db	1452	GAGAGCAAGAACGATGAGGAGTTTCAACAGATTCACCTGTGATGTTCACTGAGAGTAC	1511	PUBMED	Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
Qy	566	GLNVALGIMETLEUASPMETTYRHISARGVALPHEASP-----ARGILEGLUSERME	583	REFERENCE	RIKEN integrated sequence analysis (RISA) system-384-format
Db	1512	CAGAAAAGCTGCTAGAGAGATGACATCTGCGTCTGATCAAAAACGCAAAAATACGTG	1571	AUTHORS	sequencing pipeline with 384 multicapillary sequencer
Qy	584	ALAGLYGLUHIYVALTPHANPHEALAAPRHEGLINTHRANLEUGLYLEILEARGVAL	603	TITLE	Genome Res. 10 (11), 1757-1771 (2000)
Db	1572	GTTGGAGACTCTATTTGGCAATTTTGGCGATTTTCATGACGTCAACAGTACCGACGAGATG	1631	JOURNAL	4 The RIKEN Genome Exploration Research Group Phase II Team and the
Qy	604	AAPGLYASRLYHLYAGLYVALPHEATHRARGAPARGLYSERPROYALALALAHLSER	623	REFERENCE	PANTOM Consortium.
Db	1632	CTGGGGAAATTAAGAGGATCTTCACTCGGACAGACCAACAAAGATGAGGCTTCTT	1691	AUTHORS	Functional annotation of a full-length mouse cDNA collection
Qy	624	LEUARGALARGTRPTHRSERIE 631		TITLE	Nature 409, 685-690 (2001)
Db	1692	TTGCCGAGAGATGATCTGMAAGATT 1715		JOURNAL	5 The PANTOM Consortium and the RIKEN Genome Exploration Research
RESULT 4				REFERENCE	Group Phase I & II Team.
AK041058	2274 bp	RNA	linear	AUTHORS	Analysis of the mouse transcriptome based on functional annotation
LOCUS				TITLE	of 60,770 full-length cDNAs
DEFINITION				JOURNAL	Nature 420, 563-573 (2002)
				REFERENCE	6 (bases 1 to 2274)
				AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
					Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
					Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
					Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
					Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M.,
					Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
					Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohsato, N.,
					Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
					Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
					Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
					Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasumishi, A.,
					Muramatsu, M. and Hayashizaki, Y.
					Direct Submission
					Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
					Physical and Chemical Research (RIKEN), Laboratory for Genome
					Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
					RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
					Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp,
					URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
					Fax: 81-45-503-9216)
					CDNA library was prepared and sequenced in Mouse Genome
					Encyclopedia Project of Genome Exploration Research Group in Riken
					Genomic Sciences Center and Genome Science Laboratory in Riken.
					Division of Experimental Animal Research in Riken contributed to
					prepare mouse tissues.
					Please visit our web site for further details.
					URL: http://genome.gsc.riken.jp/
					URL: http://phantom.gsc.riken.jp/
					Location/Qualifiers
					1. 2274
					/organism="Mus musculus"
ACCESSION	AK041058				
VERSION	AK041058.1	GI:26088231			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				

FEATURES

source

```

/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A530072005"
/db_xref="taxon:10090"
/clone="A530072005"
/bex="male"
/tissue_type="aorta and vein"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
14. 1778
/note="beta-glucuronidase structural (MGI:95874,
GB|NM_010368, evidence: BLASTN, 99%, match=733)
putative"
2260. 2265
/note="putative"
2274
polya_signal
2274
polya_site
/note="putative"

ORIGIN

Alignment Scores:
Pred. No.: 1,73e-119 Length: 2274
Score: 1168.50 Matches: 249
Percent Similarity: 55.76% Conservative: 90
Best Local Similarity: 40.95% Mismatches: 179
Query Match: 34.84% Indels: 90
DB: Gaps: 13

US-10-757-093-4 (1-634) x AK041058 (1-2274)

QY 42 ProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu 61
DB 101 CCGAAGGAGAGCCCGCTCGCGAGCTCCAGCGCTGACGACTGTGGCACTTCGCGCC 160
QY 62 -----AlaSerGlyLeuAsnAspThrAlaGlnProTrp---ThraAlaProLeuPro 77
DB 161 GACCTCTCGAACAACCGGCTGCGAGGTTTGAGACGAAATGTAACCGGCGGCTTACGG 220
QY 78 LysGly-----LeuGlyCysProValProAlaSerTyraAsnAspIlePheIleSer 94
DB 221 GAGTCGGGCGCCAGCTTGGACATGCTGTCTCTTACGTTCAATGACATCACCAAGAA 280
QY 95 ArgGlnIleHisAspHisValGlyTrpValTyrrTyrrGlnArgGluValIleValProLys 114
DB 281 GCAGCCCTTGGAGCTTATTTGGCTGGTGTGTATGAACGGAGAACATCCCGCACCG 340
QY 115 GlyTrpSerGlnGlu-----ArgTyrrLeuValArgAlaGluSerAlaThrHis 131
DB 341 CGATGACCCCAATATCCGACATGAGAGTGTGTGAGATCAACAGTGC---CATTTAT 397
QY 132 GlyArgIleTyrrValAsnAsnAspGluValAlaGluHisValGlyTyrrThrProPhe 151
DB 398 TATGCAATTGTGTATCCCA----- 418
QY 152 GluAlaAspValThrGluLeuValAlaProGlyGluLys-PheArgLeuThrIleGlyVal 171
DB 419 -----GGGTTACTTTGTCCAGGACACAAAGCTT----- 446
QY 171 LeuAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThyAsnAlaThr 191
DB 446 ----- 446
QY 191 rGlyLysArgIleGlnThrTyrrGlnHisAspPheTyraAsnTyrrAlaGlyLeuAlaArgSe 211
DB 447 -----GACTTCTCAACTATGCGGAGCTGCATCGATC 478
QY 211 rIleTrpLeuTyrrSerValProGlnGlnHisIleGlnAspIleThrValValThrAspHis 231
DB 479 TGTGTCCTCTAATACCACTTACCACTGATGATATCACTGTGATCACTAATGT 538
QY 231 LaSpGlyAspAsnGlyLeuIleAsnTyrrGluValAlaAsnGlnThrThrGlyVal 251
DB 539 GGAGCAAGACATGGGCTGTGTGACCTAATCGATTTCGTGCGAGGACAGTGAACATTCCA 598

```

```

QY 251 nIleGlnIleSerValIleAspGluAspGlyValAlaIleValAlaIleAspGlyValAla 271
DB 599 GCTAAGAAAGTGAACATTTTGATGATGAGGCTGAAGAGCTGTGGCCATGGGACAGGAACA 658
QY 271 nGlyThrValThrIleProSerValLysLeuTrpGlnPro-----GlyVal 286
DB 659 GGGTCAACTTCAGTTCCAGTCCAGGCCAACCTGTGGTGGCTTACCTGATGATGAGCATCC 718
QY 286 alaTyrrLeuTyrrGlnLeuGluValAsnIleValGlySerSerGlyAspValValAspHis 306
DB 719 AGCTTACATGACTCTCTTGAGAGTGAAGTGA---ACAAGTGAAGTGTGATGACTACTA 775
QY 306 rTyrrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAs 326
DB 776 CTACACCTTCTCTATGCGGATTCGAACAGTGTGCTTCAAAAGAGAAAGTTCCTCTATAA 835
QY 326 nGlyLysProPheTyrrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyVal 346
DB 836 CCGGAAGCCCTCTAATTCGAAGGGTCAATAGACGAGAGATTCAGATATCCGAGGGA 895
QY 346 sGlyHisAspProAlaTyrrMetValHisAspPheGlnLeuMetLysTrpIleGlyValAs 366
DB 896 AGCTTCGACCTGGCCGCTGCTGTGAAGATTTCAACCTGCTCCTGTGGCTGGGCGA 955
QY 366 nSerPheArgThrSerHisTyrrProTyrrAlaGluGluValMetAspPheAlaAspArgAs 386
DB 956 TTCCTTCGTACAGCACTATCCCTACAGAGAGGTATCTTACGCTGTGACCGATA 1015
QY 386 nGlyIleValIleIleAspGluThrProAlaValGlyLeuAsnIleAlaLeuMetGlyVal 406
DB 1016 CCGGATTTGTGATGATGATGATGTCGCGTGTGGCATTTGTCTA----- 1061
QY 406 lSerGluSerGlyAlaProGlnThrPheThrProAspAlaIleAsnAspLysThrGlnGln 426
DB 1062 -----CTCCAGATTTGGCAACAGATCACTTCGGCAC----- 1094
QY 426 uAlaHisGlyAlaIleAlaArgGluLeuIleAlaArgAspLysAsnHisAlaSerValVal 446
DB 1095 -----CACCTAAGCGATGATGAGAGCTGTGTGCGGCGCAAAATATCACCCGTGGTGT 1150
QY 446 lMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGluTyrrPheGln 466
DB 1151 GATGTGTGTGTGGCCAAAGACCTTCTCTGTCTGAAACCGCGCATATTACTTTAA 1210
QY 466 uProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnVal 486
DB 1264 1211 GAGCGTGAACCCACACCAACAAAGCCCTGACCTCACCGCTCCCGTACCTT-----GT 1264
QY 486 lGlyThrAlaThrTyrrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAs 506
DB 1265 GAGCAACGCGCAATATGATGACAGCTGGGCGGCCCGTACGTGAGATGTTATCTGTGTA 1324
QY 506 nArgTyrrPheGlyTyrrTyrrSerGlnThrGlyLysPheGluGluAlaIleGln 526
DB 1325 CAGCTACTTTTCTTGATCATGACTATGAGGATTTGAGAGTATTCAGCCACAGCTGAA 1384
QY 526 uLysGluLeuHisGlyTyrrGlnGluLysPheHisAspProIleValMetThrGluTyrrGln 546
DB 1385 TAGCAGCTTGAAGACTGTATTAAGACGATCAGAAAGCGATTAACAGAGGAGTATGG 1444
QY 546 yAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGlyGluPheGln 566
DB 1445 AGCAGACGCGAATCCAGAGGATCCAGAGAGACCCGCTCGCATGTTCAGTGAAGAGTACCA 1504
QY 566 nValGlnMetLeuAspMetTyrrHisArgValPheAsp-----ArgIleGluSerMetAl 584
DB 1505 GAAGCTGTCTGAGAAATTACATTCAGTTTGTGATCAAGAAAGTAAAGATACGTGCT 1564
QY 584 aGlyGluHisValTyrrAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAs 604
DB 1565 CCGAAGACATCATCTGGAATTTGCGGACTTCAAGCAAGACAGTCAACCGCTGAGATAT 1624
QY 604 rGlyAsnLysLysGlyValPheThrArgAspArgLysProLysValAlaIleHisSerIle 624

```


Db 1625 CCGAACAAGAGGGGATCTTCATCGCCAGAGACAGCCCAAACTTCGGCTTATTTT 1684
 QY 624 uArgAlaArgTPrThSerIle 631
 Db 1685 GCGAGAGAGTACTGAGAGATT 1706

RESULT 5
 CD014093 1853 bp mRNA linear EST 21-OCT-2003
 LOCUS CD014093
 DEFINITION Single gene library Homo sapiens cDNA, mRNA sequence.
 ACCESSION CD014093
 VERSION CD014093.1 GI:37777622
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1853)
 Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R.,
 Au-Yang, J., and Stuve, L.L.
 PCR isolation and cloning of novel splice variant mRNAs from known
 drug target genes
 Genomics 83 (4), 566-571 (2004)
 CONTACT: Jin, P.
 Jincyte Corporation
 3160 Porter Drive, Palo Alto, CA 94304, USA
 Tel: 650 621 8639
 Fax: 650 621 8965
 Email: pjin@jincyte.com.
 Location/Qualifiers
 1..1853
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="Single gene library"
 /note="Vector: pDrive Cloning Vector; RT-PCR was performed
 using gene-specific primers flanking the open-reading
 frame. PCR products were subcloned into pDrive Cloning
 Vector and sequenced completely using M13 forward and
 reverse primers. Sequencing gaps were closed by
 re-sequencing using primers flanking the gapped areas."

ORIGIN

Alignment Scores:
 Pred. No.: 8,53e-106 Length: 1853
 Score: 1045.50 Matches: 227
 Percent Similarity: 50.90% Conservative: 84
 Best Local Similarity: 37.15% Mismatches: 159
 Query Match: 31.17% Indels: 141
 DB: 6 Gaps: 13

US-10-757-093-4 (1-634) x CD014093 (1-1853)

QY 42 ProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu 61
 Db 101 CCCGAGAGAGCCCGCGCGGAGTCCAGAGAGCTGAGCGCCTCGAGCTTC----- 154
 QY 62 AlaSerGlyLeuAsnAspThr-----AlaGlnProTrpThraLPro 75
 Db 155 CCGCGGAGCTTCTCTGACAAACGAGCGCGGCGCTTCGAGAGACAGTGTACCGCGCGCG 214
 QY 76 LeuProLysGly-----LeuGlnCysProValProAlaSerTyrrhAsnIlePhe 92
 Db 215 CTGTGGAGATCAGGCCCAACCGGTGACATGCCATTCCTCCAGCTTCATATACATCAGC 274
 QY 93 IleSerArgGluIleHisAspHisValGlyTrpValTyrrGlnArgGluValIleVal 112
 Db 275 CAGGACTGCGCTCTGCGGATTTTGTGCGCTGGGTGTGTGTAAGACGAGAGTGTCTCG 334
 QY 113 ProLysGlyTrpSerGlnGlu-----ArgTyrrLeuValArgAlaGlnSerAlaThr 129
 Db 335 CCGGAGCGATGTGACCGACGACTCGCGCACAGAGTGTGTCTGAGATTGGCAGTGC--- 391

QY 130 HisHISGlyArgIleTyrrValAsnAspArgLeuValAlaGlnHisValGlyGlyTrpThr 149
 Db 391 ----- 391
 QY 150 ProPheGlnAlaAspValThrGlnLeuValAlaProGlyGluLysPheArgLeuThrIle 169
 Db 391 ----- 391
 QY 170 GlyValAsnAsnGlnLeuThrHisGlnThrIleProProGlyLysIleThrThrGlyAsn 189
 Db 391 ----- 391
 QY 190 AlaThrGlnLysArgIleGlnThrTyrrGlnHisAspPheAsnTyrrAlaGlyLeuAla 209
 Db 392 -----CATTCCTATCC----- 403
 QY 210 ArgSerIleTrpLeuTyrrSerValProGlnGlnHisIleGlnAspIleThrValAlaThr 229
 Db 404 -----ATCGTT----- 409
 QY 230 AspValAspGlyAspAsnGlyLeuIleAsnTyrrGluValGluValAlaAsnGlnThrThr 249
 Db 410 -----GGCTGTGGATTCACAGATCTCTGTCAAGGCGACGTAACTCG 451
 QY 250 GlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGly 269
 Db 452 TTCAGTTCGAAGTGCCTCTTTGCAATGCAGAAACAAAGTGTGCGCAATGGAGCTGG 511
 QY 270 AlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnPro----- 284
 Db 512 ACCGAGGCCCACTTAAGTGTCCAGAGTGCACCTCTGTGGCCGTAAGCTGACGAA 571
 QY 285 GlyAlaAlaTyrrLeuTyrrGlnLeuGlnValAsnIleValGly-----SerSerGlyAsp 302
 Db 572 CCGCCCTCATCTGTATTCATGTAGAGGAGTGCAGTGCACGACAGACGTCACTGGGCGCT 631
 QY 303 ValIleAspThrTyrrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGln 322
 Db 632 GTGTCTGACTTTTACACACTCCCTGTGGGAGATCCGACGTGTGCTGCACCAAGCCG 691
 QY 323 PheLeuIleAsnGlyLysProPheTyrrPheThrGlyPheGlyLysHisGlnAspThrAla 342
 Db 692 TTCCTCATCAATGGGAACCTTTCTATTTCCACGGTGTCAACAGATATGAGATGGCGAC 751
 QY 343 ValArgGlyLysGlyHisAspProAlaTyrrMetValHisAspPheGlnLeuMetLysTrp 362
 Db 752 ATCCGAGGGAAGGCGCTTCGACTGCGCGCTGTGTGAAGGACTTCACCTGCTGCGCTGG 811
 QY 363 IleGlyAlaAsnSerPheAspThrSerHisTyrrProTyrrAlaGlnGluValMetAspPhe 382
 Db 812 CTGTGGGCCAAGCCTTCCGATCCAGCCACTACCCCTATGTCAAGAGAAATGTAGTGCAGATG 871
 QY 383 AlaAspArgAsnGlyIleValValIleAspGlnThrProAlaValGlyLeuAsnIleAla 402
 Db 872 TGTGAACCGTATAGGATGTGTGTATCATGATAGTGTCCCGGGGTGGCGCTGGCGCTG--- 928
 QY 403 LeuMetGlyValSerGlnSerGlyAlaProGlnThrPheThrProAspAlaIleAsnAsp 422
 Db 929 -----CCGCGATTCCTC-----AACAC 946
 QY 423 LysThrGlnGlnAlaHisLysGlnAlaIleAspGlnLeuIleAlaArgAspLysAsnHis 442
 Db 947 GTTCTCTGCATCACCACTGCAGTGTATGGAAGAAAGTGTGTGCTGGGACAAAGAAC 1006
 QY 443 AlaSerValAlaMetTrpSerIleAlaAsnGluProAlaSerHisGlnAspGlyAlaArg 462
 Db 1007 CCGGAGTGTGATGTGTGTGTGTGTGCAACAGAGCTGTGGTCCACTTAAGATCTGTGGC 1066
 QY 463 GlnTyrrPheGlnProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThr 482
 Db 1067 TACTACTTGAAGATGTATGTGCTGCACCAATCTTGTGACCCCTCGGCGCTGTGACC 1126


```

Qy 549 ThrLeuAlaGlyLeuHisSerIleLeuGlyProTyrSerGluGluPheGlnValGln 568
Db 910 ACATTGACAGGGTTTACACGATCCACCTGATGTTACCTGAAGAGTACGAAAGT 969
Qy 569 MetLeuAspMetCysHisArgValPheAsp-----ArgIleGluSerMetAlaGlyGlu 586
Db 970 CTGCTAGACACAGTACATCTGGCTCTGATCAAAAAACGCAAAAATACGCTGTGAGAG 1029
Qy 587 HisValTyrAspPheAlaAspPheGlnTyrAsnLeuGlyIleIleArgValAspGlyAsn 606
Db 1030 CTCATTGGAATTTCCCGATTTCATGACTGAACACTCCGACGAGAGTCTGGGGAAAT 1089
Qy 607 LysLysGlyValPheThrArgAspArgLysProLysAlaAlaAlaHisSerLeuArgAla 626
Db 1090 AAAAAGGAGATCTTCACTCGGACAGACAAACAAAAGTCCACGTTCTTTGCCAGAG 1149
Qy 627 ArgTyrThrSerIle 631
Db 1150 AGTACTGGAGAGATT 1164

RESULT 7
CD503076 1124 bp mRNA linear EST 12-JUN-2003
LOCUS CDA60-C07.x1d-c SHGC-CDA Gaerosteus aculeatus cDNA clone
DEFINITION CDA60-C07 5', mRNA sequence.
ACCESSION CD503076
VERSION CD503076.1 GI:31432977
KEYWORDS EST.
SOURCE Gaerosteus aculeatus (three spined stickleback)
ORGANISM Gaerosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gaerosteiformes;
Gaerosteidae; Gaerosteus.
1 (bases 1 to 1124)
Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,
Schmutz, J. and Myers, R.M.
Expressed sequence tags from Gaerosteus aculeatus
Unpublished (2003)
Contact: Kingsley, DM
HIMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@cmgm.stanford.edu
Plate: 60
High quality sequence stop: 782.
Location/Qualifiers
1..1124
/organism="Gaerosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/cclone="CDA60-C07"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_1ib="SHGC-CDA"
/notes="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI
(5' adaptor); Site 2: XhoI (3' linker primer). The mixed
organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dt sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."

```

```

ORIGIN
Alignment Scores:
Pred. No.: 7,26e-79 Length: 1124
Score: 803.00 Matches: 164
Percent Similarity: 60.74% Conservative: 48
Best Local Similarity: 46.99% Mismatches: 115
Query Match: 23.94% Indels: 23
DB: 6 Gaps: 5

US-10-757-093-4 (1-634) x CD503076 (1-1124)
Qy 284 ProGlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspVal 303
Db 8 CGAACCAGGATTAACCTTACCTTTGAGAGTGTCTTACTGCGCTGAGAGATCGCT 67
Qy 304 Val--AspThrTyrAsnLeuAlaThrGlyValAlaGlyThrValLysValAlaGlySerGln 322
Db 68 TCCAACGACGTGTACACTTACCAGTCCGATCCGACCGCTGACCTTACCAAGACCCAG 127
Qy 323 PheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAla 342
Db 128 TTCCTCATCAACAAAAGCCCTTCTACTTCCACGAGTAATTAACACAGAGACTGTAT 187
Qy 343 ValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTyr 362
Db 188 ATTCCAGGCAAAAGCCCTGAGCTGCCCCCTCATGCTGAAGACTTTAATTGAAAGTGG 247
Qy 363 IleGlyAlaAsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPhe 382
Db 248 TTGGGGGGCAACTCGTTCCGACACAGCCACTACCTTATGACAGAGAGATCTCGCAGATG 307
Qy 383 AlaAspArgAsnGlyIleValValIleAspGluThrProAlaValGlyLeu---AsnIle 401
Db 308 TGTGACCGCCAGGACATCTGTGTGATGACGAGTGGCCGCGCTGCATAAAGCAATT 367
Qy 402 AlaLeuMetGlyValSerGluSerGlyAlaProGlnThrPheThrProAspAlaIleAsn 421
Db 368 CGCAGTTTGG----- 379
Qy 422 AspLysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsn 441
Db 380 AACGCTCTTAACCATCAGTACCTGTCGTCATGACGACGACGACTGTGATCGTGGGCAAGAAC 439
Qy 442 HisAlaSerValValMetTyrSerIleAlaAsnGluThrProAlaSerHisGluAspGlyAla 461
Db 440 CATCCCTGTGTGTATGTCATGTCAGTGGCCATGACCGGCTGCACAGATGCTCTGTGT 499
Qy 462 ArgGluTyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIle 481
Db 500 GATTACTATTTCAAACCTTGAATAAACAATACCAAGATTGATTCACCGGCCCTGTC 559
Qy 482 ThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAsp 501
Db 560 ACTTT-----ATCACACAGAGTAATATGACGAGGATAAAGGGGCTCCCTACGCGAC 613
Qy 502 ValSerCysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluGluAla 521
Db 614 GTCATCTGGCTAAACAGTTACTTCTCTGTGACATGACCCGGGCGCACCCGAGCATGTC 673
Qy 522 GlnAlaIleLeuGluLysGluLeuHisGlyTyrProlGlnGlyPheHisArgProIleVal 541
Db 674 CCCATCCAGCTCAACACTGATGTTGAGAACTGTACGGAAGTACCAAGAACCAATCATATC 733
Qy 542 MetThrGluTyrGlyAlaAspThrLeuAspThrLeuHisSerIleLeuGlyLeuProTyr 561
Db 734 CAGAGGAGATTAAGGAGCGGATGCGGGCGCTTCAAGATGATCCACCCGATGTTT 793
Qy 562 SerGluGluPheGlnValGlnMetLeuAspMetCysHisArgValPheAsp-----Arg 579
Db 794 ACTGAGAGTACAGAGAGTATGCTTGCAGAGCTTCCACACAGATGTTTCCACCAAGAAAG 853
Qy 580 IleGluSerMetAlaGlyLysHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599

```

Db 854 AAGCAGTACGTACCGCGAACTCAGTCTGCAAGCTTGCAGATTCATGACCCACAGAGG 913
 Qy 600 TlellArVAlAspGlyAsnLysGlyValPheThrArgAspArgLysPhePolyAla 619
 Db 914 ATATCGCGTGTGGGAAACAAAGAGGTCTCTCAGAGGCA-AGGAGCCCAAGCAG 972
 Qy 620 AlaAlaHisSerLeuArgAlaArgTTP 628
 Db 973 CAGCATCATCTGAAGGAGAGTACTGG 999
 RESULT 8
 LOCUS BX363460 1051 bp mRNA linear EST 08-APR-2004
 DEFINITION BX363460 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
 Homo sapiens cDNA clone CSDDL001YM21 5-PRIME, mRNA sequence.
 ACCESSION BX363460 GI:46291531
 VERSION BX363460.2
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (Drees 1 to 1051)
 Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On May 5, 2003 this sequence version replaced gi:30372720.
 CONTACT: Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: sequef@genoscope.cns.fr, web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 4169.r
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?c=CSDDL001AG11QPlc=4169.r.

FEATURES

source

Location/Qualifiers

1..1051
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSDDL001YM21"
 /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
 /cell_line="RAMOS CELL LINE"
 /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 2,72e-78 length: 1051
 Score: 797.50 Matches: 159
 Percent Similarity: 59.50% Conservative: 57
 Best Local Similarity: 43.80% Mismatches: 121
 Query Match: 23.78% Indels: 26
 DB: 5 Gaps: 6

US-10-757-093-4 (1-634) x BX363460 (1-1051)

Qy 236 GlyLeuIleAsnTyrgluValAlaAsnGlnThrThrGlyGlnIleGlnIleSer 255
 Db 1 GGGCTGTGATTAACAGATCTCTGTCAAGGGCAGTAACCTGTTCAAGTTGAAGGCCG 60
 Qy 256 ValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyValAglnglyThrValThr 275
 Db 61 CTTTGGATGACAGAAACAAAGTCGTGGCAATGGAGCTGGGACCCAGGGCCAACTTAAG 120

Qy 276 IleProSerValLysLeuTTPGlnPro-----GlyAlaAlaIleuTyr 290
 Db 121 GGCCAGCGGTGAGGCTCTGGGGCCGTACCGATGCACGAAGCCCTCTATCTGAT 180
 Qy 291 GlnLeuGlnValAsnIleValGly-----SerSerGlyAspValValAspThrTyrAsn 308
 Db 181 TCATTGGAGGTGACGTGACCTGACACAGACGTGAGGGCCGTGTCTTCAATTCACACA 240
 Qy 309 LeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLys 328
 Db 241 CTCCTGTGGAGATCCGACCTGTGGCTGTACCAAGACCACTTCTCTCAATGGGAA 300
 Qy 329 ProPheTyrPheThrGlyPheGlyLysHisGlyAspThrAlaValArgGlyLysGlyHis 348
 Db 301 CTTTCTATTTCAGAGTGTCAACAAGCATGAGATGCGGACATCGGAGGAGGAGGCTTC 360
 Qy 349 AspProAlaTyrMetValHisAspPheGlnLeuMetLysTTPILeglyAlaAsnSerPhe 368
 Db 361 GACTGGCCGCTGCTGGTGAAGGACTTCAACTGCTTGGCTTGGTCCAAACGCTTTC 420
 Qy 369 ArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArgAsnGlyLys 388
 Db 421 CGTACCAAGCCACTACCTTATGCAAGAGAGATGATGACATGTGACCGCTATGGGATT 480
 Qy 389 ValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeuMetGlyValSerGlu 408
 Db 481 GTGGTCATGATGAGTGTCCGCGGTGGCTGGCGCTG----- 519
 Qy 409 SerGlyAlaProGlnThrPheThrProAspAlaIleAsnAspLysThrGlnIleuAlaHis 428
 Db 520 -----CCGACAGTCTTC-----AACACGGTTCCTTGATCACCAC 555
 Qy 429 LysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTyr 448
 Db 556 ATGCAGGTATGAGACAAAGTGTGCTGAGGACAAAGACACCCGCGCTGTGATGTGG 615
 Qy 449 SerIleAlaAsnGluProAlaSerHisGlyAspGlyAlaArgGluTyrPheGluProLeu 468
 Db 616 TCTGTGCCACAGAGCTGCTGCCACCTAGATCTGTGCTACTTGAAGATGTG 675
 Qy 469 ThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThr 488
 Db 676 ATCGCTCACACCAAAATCCCTTGAACCCCTCCGCGCTGTATACCTT-----GTGAGCAAC 729
 Qy 489 AlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSerCysAlleAsnArgTyr 508
 Db 730 TCTAATCTATGCACGACAGAGGGGCTCGGTATGTGATGTATCTTGTGAACAGCTAC 789
 Qy 509 PheGlyTTPtyrSerGlnThrGlyAspLeuGluGluAlaGluAlaLeuGluLysGlu 528
 Db 790 TACTCTGTGATACAGACTACGAGGACCTGAGGTGATTACACTGCAGCTGCACCCAG 849
 Qy 529 LeuHisGlyTTPGlnGluLysPheHisArgProIleValMetThrGlyTyrGlyAlaAsp 548
 Db 850 TTTGAGAACTGGATATAGATATCAGAACCCATTATCAGACGAGATATGAGACAGTA 909
 Qy 549 ThrLeu-AlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGlnGluPheGluValGlu 568
 Db 910 ACGATTGAGGGGTTTCCACAGATTCACACTCTGATGTTCACTGAGAGTAACAGAAAR 969
 Qy 568 nMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetValGlyLysHis-- 587
 Db 970 TTYTCTAGARCAGTACCATCTGGTCTGTGATTAATAAAMGAAATATATGTGTGGRCT 1029
 Qy 588 -ValTTP 589
 Db 1030 CATTTGG 1036
 RESULT 9
 LOCUS CL486845/c 921 bp DNA linear GSS 01-APR-2004
 DEFINITION SAIL_443_B06.v2 SAIL Collection Arabidopsis thaliana genomic clone

FEATURES	Location/Qualifiers
source	1. .921
	/organism="Arabidopsis thaliana"
	/mol_type="genomic DNA"
	/ecotype="Columbia"
	/db_xref="taxon:3702"
	/clone="SAIL 443 B06.v2"
	/clone_1bp="SAIL Collection"
	/note="T7-DNA left border sequences were isolated using a modified Tail-Pick strategy"

[illegible]

a 50 bp linker primer containing an oligo dt sequence preceded by a synthetic XhoI site. 5 prime adaptors were used containing an EcoRI cohesive end. The finished cDNAs were inserted in to the ZAP express vector unidirectionally in the sense orientation with respect to the lacZ promoter of pBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual pBK-CMV phagemid clones for EST sequencing."

ORIGIN

Alignment Scores:

Pred. No.:	3.34e-75	Length:	1138
Score:	770.50	Matches:	166
Percent Similarity:	61.56%	Conservative:	47
Best Local Similarity:	47.98%	Mismatches:	111
Query Match:	22.97%	Indels:	25
DB:	6	Gaps:	5

US-10-757-093-4 (1-634) x CD503098 (1-1138)

```

QY 284 ProGlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspVal 303
DB 8 CGAACCAGGCTTACTCTTCTGAGGCTTCTTACTGCTGCTGAAGGATCTGCT 67
QY 304 Val--AspThrTyrAsnLeuAlaThrGlyValAlaGlyValAlaGlySerGln 322
DB 68 TCCACGACGCTGACCTCTACCGAGTCCGACGCGTCCGACGCTTACCGACGAC 127
QY 323 PheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHISGlnAspThrAla 342
DB 128 TTCCTCATCAAAAGCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 187
QY 343 ValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnMetLysTyr 362
DB 188 ATTCGAGGCAAGGCTGCGAGTCCCTCATGCTGAGAGCTTACTTACTTATGAGAGT 247
QY 363 IleGlyIleAsnSerPheArgThrSerHisTyrProTyrAlaGlnGluValMetAspPhe 382
DB 248 TTGGGGGCGCAAGCTGTTCCGACGACGACCTTATGACGAGGAGATCTGACAGATG 307
QY 383 AlaAspArgAsnGlyIleValAlaIleAspGlnThrProAlaValGlyLeu--AsnIle 401
DB 308 TGTGACCGCATGCGCATCTGTGTATGACGAGCGCGGCGTGGGCAATAAAGACAT 367
QY 402 AlaLeuMetGlyValSerGlyValAlaProGlnThrPheThrProAspAlaIleAsn 421
DB 368 CGCAGTTTGA----- 379
QY 422 AspLysThrGlnGluAlaHisLysGlnAlaIleArgGlnLeuIleAlaArgAspLysAsn 441
DB 380 AACGCCCTTAAACCATCACTGCTGCTGATGACGACGCTGCTGCTGCTGCTGCT 439
QY 442 HisAlaSerValValMetTyrSerIleAlaAsnGluProAlaSerHisGlnAspGlyAla 461
DB 440 CATCCCTCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 499
QY 462 ArgGlnTyrPheGlnProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIle 481
DB 500 GATTACTATTCAAAACCTTGATMAACATACCAAGAAATGGATCAACCGCGCCG 559
QY 482 ThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerSerPheLeuAsp 501
DB 560 ACTTTT-----ATCACAGACAGTACTATGACGAGGATMAAGGGCTCTTACGTGAC 613
QY 502 ValSerCysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGlnGluAla 521
DB 614 GTGATCTGGGTAAACAGTACTCTCTGCTGATCAAGACCGGACACCCGAGGTATC 673
QY 522 GluAlaAlaLeuGlnLysGlnLeuHisGlyTyrGlnGlnLysPheHisArgProIleVal 541
DB 674 CCATCAGCTCAACCACTGATTGAGAACTGTATCGGAAGTAACAGAAACCATATC 733

```

```

QY 542 MetThrGlnTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyr 561
DB 734 CAGAGCGAATACAGAGCGATGCGTCCGGGCTTCAAGTATCCACCCGATGTT 793
QY 562 SerGlnGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAsp-----Arg 579
DB 794 ACTGAGAGTACCAAGAAAGTTAGTCTGACAGCTTACCAAGCTGTTCCACAGAAAGA 853
QY 580 IleGluSerMetAlaGlyGlnHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599
DB 854 AG-CAGTACGTCAATCGCGAATCATCTGGAATTTGACGACTTATGACACAGAG 912
QY 600 IleIleArgValAlaAspIleValLeuLysGlyValPheThrArgAspArgTyrPheVal 619
DB 913 ATCATCGGCTGCTGGGAAACAGAG-GGTGCTTACAGCAGGAAAGGACCAAGCA- 970
QY 620 AlaAlaHisSerLeuArg 625
DB 971 GCAGATTCATCTTAAG 988

```

RESULT 11

BX745933/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 857
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TGAS066d19"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5' end of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into PCS107 with Host: Escherichia coli XL1-blue.
Location/Qualifiers

ORIGIN

Alignment Scores:

Pred. No.:	1.9e-70	Length:	857
Score:	726.50 <td>Matches:</td> <td>144</td>	Matches:	144
Percent Similarity:	63.95%	Conservative:	44
Best Local Similarity:	48.98%	Mismatches:	85

Query Match:	21.66%	Indels:	21
DB:	5	Gaps:	5
US-10-757-093-4 (1-634) x BK745933 (1-857)			
Qy	283 GlnProGlyAlaAlaIaTYLeuTYrGlnLeuGlnValAenIleValGlySerSerGlyAsp	302	
Dy	849 GAGCTCGA-----TACTGTACTCAATTGAGAGGTATGTTATGTCGCAAGATGGAAAT	796	
Qy	303 -----ValValAspThrTYrAsnLeuAlaThrGlyValArgThrValLysValAlaGly	320	
Dy	795 GGGTCAGTGAAGAACAATTACATTTTGGCTGGGAAATAGAAACTGTCAAGCTCTCGA	736	
Qy	321 SerGlnPheLeuIleAsnGlyLysProPheTYrPheThrGlyPheGlyLysIleGluAsp	340	
Dy	735 GACCAATTTCCATCATCAATGAAACCAATTCACTTCATGGTGTCACAAACATGAGAC	676	
Qy	341 ThrAlaValArgGlyLysGlyVNIAspProAlaTYrMetValIleAspPheGlnLeuMet	360	
Dy	675 TATGATGTCAGAGAAAGAAAGACTAGACTGTGCACATATTGTAAAGATTTTATCTGTG	616	
Qy	361 LysTrpIleGlyAlaAsnSerPheArgThrSerHisTYrProTYrAlaGluGluValMet	380	
Dy	615 AAGGGCTGGTGGTAACTCTTCGCGACAGCATTTATCTTATGACGAAGAATCATG	556	
Qy	381 AspPheAlaAspArgAsnGlyTYrLeValIleAspGluThrProAlaValGlyLeuAsn	400	
Dy	555 GACCTTGTGTAAATATGACATGTGTGTGATGTGATGATGATGCCCTGAGTGGTAAAA	496	
Qy	401 IleAlaLeuMetGlyValSerGluSerGlyAlaProGlnThrPheThrProAspAlaIle	420	
Dy	495 TAC-----CCGAAAGTTTGGGAAACCAATCTTTA	466	
Qy	421 AsnAspLysThrGlnGluAlaHisIleGlnAlaIleArgGluLeuIleAlaArgAspLys	440	
Dy	465 AACCAC-----CATTTAATAGTTATGAGAGAGTTGGTGGCAGGATAA	421	
Qy	441 AsnHisAlaSerValValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGly	460	
Dy	420 AACCGGCTTCGCTGATGATGGCTGTGGCAATGAAACCAAGCATCCAGCTCCCTGTG	361	
Qy	461 AlaArgGluTYrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgPro	480	
Dy	360 GCTGGGTATTTTAAACCGTGATGTTGCTTACCTAAGCAACTTGACCCCACTGGACCA	301	
Qy	481 IleThrPheAlaAsnValGlyThrAlaThrTYrGlnLeuAspArgIleSerAspLeuPhe	500	
Dy	300 GTGACATAT-----GTTTCCATGCTTAACATGAAACATGACCAAGTGCACCTATG	247	
Qy	501 AspValSerGlyIleAsnArgTYrPheGlyTYrTYrSerGlnThrGlyAspLeuGluGlu	520	
Dy	246 GATGTAATTTGTGTAAACAGTTACTTTCTTGATCATGATGACGGGACACTGGAAAT	187	
Qy	521 AlaGluAlaIleLeuGluLysGluLeuHisGlyTYrGlnGlnGluLysPheHisArgProIle	540	
Dy	186 ATCCAACTCCAACTCAATGATCAGTTTGACAAATGATGAGAAATATCAAAAGCTTAG	127	
Qy	541 ValMetThrGluTYrGlyValAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuPro	560	
Dy	126 ATACAGATGAATATGAGACAGATTCATTTCCAGGGTTTCAACAGTGAACCCCATTAATG	67	
Qy	561 TrpSerGluGluPheGlnValAlaGlnMetLeuAspMetTYrHis.574		
Dy	66 TTCACTGAAGAATACCAAGAGTTGTGTAAAGAAACTACAT	25	
RESULT 12			
LOCUS	CL679241	811 bp	DNA linear
DEFINITION	PRJ0125C C10.2 - PRJ0125C_BR (811) Mixed stage foamed library of P		
	acificus var. California Pristionchus pacificus genomic, genomic		
	survey sequence.		
ACCESSION	CL679241		
VERSION	CL679241.1	GI:50185841	

[illegible]

Oy		255	ServallileapsgluaspghyalalilevalialalsalaserglyalagnglThrvAl	274
Dd		466	GAACTGGCTGTATCGGATCAACAGGTGTTCGACTGCAGAACAGCAGCGACTTTG	525
Oy		275	ThrilleProSerVallysleutirpGlnProglYlaalaatyLeutyrgInleucInVaI	294
Dd		526	CAA GTGGTGAATCGCACCTCTTGCCAA TC GGGGTGA AGGTATATCTTA TGA ACTGCGCTC	585
Oy		295	AsnileValglyserSergilyAspValValasphTryrAsnleuAlaethrGlyValArg	314
Dd		586	-----ACAGCCA AAGCCAGACAGAGTGTATATCTA CCGCTGCCGTGGCATCCGG	639
Oy		315	ThrvAllysValAlaaglyserGlnPheleutlleAnnglyLysProPhetyrPheThrGlyI	334
Dd		640	TCACTGGAGAGGAAGGGCGAACAAGTCTCTGATCAACCAACACGTTTCTACTTACTGCG	699
Oy		335	PhegilyLysHlsGIUASPThAlaValArGglyLySGlyHisaspProAlatYrmelVal	354
Dd		700	TTTTGGCCGCTTGAAGAATGCGGATTYYTGGCGGCAAAAGGATTCGATTAACGTGCTGATGGTG	759
Oy		355	HIsaSPhegnleumetyleTyripIeglylaaenseerPheaTyrThser	371
Dd		760	CACATTCACGATTAATGACTGATTTGGGCCCACTCTACCGTACTCG	810
RESULT 13				
LOCUS	BQ941196	906 bp	mRNA	linear EST 21-AUG-2002
DEFINITION	AGENCOURT 8741562 NIH_MGC_18 Homo sapiens CDNA clone IMAGE:6420519			
ACCESSION	BQ941196			
VERSION	BQ941196.1	GI:22356674		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	NIH-MGC http://mgc.nci.nih.gov/.			
AUTHORS	1 (bases 1 to 906)			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bs-re@mail.nih.gov Tissue Procurement: DCTD/DTP/Gazdar CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LICM2594 row: g column: 16 High quality sequence stop: 765.			
FEATURES				
SOURCE	Location/Qualifiers			
	1..906			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:6420519"			
	/tissue_type="large cell carcinoma"			
	/lab_host="DH10B (phage-resistant)"			
	/clone_id="NIH_MGC_18"			
	/note="Organ: Lung; Vector: pOTB7; Site.1: XhoI; Site.2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Nih-MGC library II RT (Life Technologies). Note: this is a			

Score:	705.50	Matches:	142
Percent Similarity:	61.44%	Conservative:	46
Best Local Similarity:	46.41%	Mismatches:	98
Query Match:	21.03%	Indels:	20
DB:	5	Gaps:	4

US-10-757-093-4 (1-634) x BQ941196 (1-906)

Qy	331	TYRPHETHGILPHHEGLIYUENHISGLIYNARPHILALVALAGSGIYLVYSGILNHSASP	Pro	350
Db	2	TATTTCCAGSGGTGTAACAAGCATGAGATGGAGCGGACCTCAAGGAAAGGCTTGACCTGG		61
Qy	351	ALAATYMEVALNHSAPRHEGLINDEUWETLUTRPILEGILYALASNSERPHETGTHR		370
Db	62	CGCGTCTGTGTAAAGSACTTCAACCTGCTGGCTGGCTGGTGGCAACGCTTTCGTAACC		122
Qy	371	SETHIETYPRTYALAGLIGLIVALELAPRHEALAPRATGASNGLYILEVALI		390
Db	122	AGCAGTACCCCTTAATGCAAGAGAGATGTATGCAATGTGTACCGCTATGGATTTGTGTC		181
Qy	391	ILASRGLIETHPRALAVAGIYUENHISNILELALUWETGLVALSERGLUSERGLY		410
Db	182	ATCGATGAGTGTCCCGCGCTGGGCTGGCGCTG-----		214
Qy	411	ALAProGILNTHrPheThrProAspAlaILEANSPRYSTRNGINGLUALNHSYGLIN		430
Db	215	---CGCGAGTTCTTC-----AACAAGTTCTCTGCATCAACCATGAC		256
Qy	431	ALALATARGLILEULILEALARGSPRYASNNHISALASERVALVAMETTRPSRIIE		450
Db	257	GTCATGAAAGATGTGTGCTGAGGACAAGAACCCCGGGCTGCGATGTGTGTGTG		316
Qy	451	ALASRGLIProAlaserISGLIYNAPRGLIALARGIYLVYRPHGLIProLeuThrASN		470
Db	317	GCCAGAGAGCTGCGTCCACCTAAGAACTGCTGCTACTACTTGAAGATGTATGCT		376
Qy	471	LEUTHRARGGLINLEUASPRTHRARGPROILETHRPHLEALASNVALGLYTHALATHR		490
Db	377	CACACCAATTCCTGGACCCCTCCGGGCTGTGACTTT-----GTAGCAACTGTAC		430
Qy	491	TYRGLINLEUASPRARGILESERASPRLEUPHEASPRVALSERCYSLILEANARGTYRPHGLY		510
Db	431	TATGACAGACAAAGGGGCTCCGTATGTGTGATGTGATCTTTTGAAACGCTACTACTCT		490
Qy	511	TRTYRSESGILNTHrGLYASPRLEUGLIGLUALAGLUALALEUGLILYSGILNHS		530
Db	491	TGGATTCACGACTACGGGACCTGGAGTTGATTTACGCTGCAAGCTGGCCACCCAACTTGGAG		550
Qy	531	GLIYTRPGLINLILYARPHENHISARGPROILEVALMETRNGIYLVYALASPRHLEU		550
Db	551	AACTGGTATTAAGATGATCAAGAGCCATTATTCAGACGAGTATGAGCGAAGAACGAT		610
Qy	551	ALAGLYLEUHHISERTILEUGLILEUPTROTTPSERGILIGLILPHNEGILVALGIMLEU		570
Db	611	GCGAGGTTTCACACGAGATCACCTGTGATGTTCACTGAAGAGTACAGAAAAAGTCTGCTA		670
Qy	571	ASPMETTYRHISARGVALPHEASPR-----ARGILEGUSERMETVALGLIUNHISVAL		588
Db	671	GAGCAGTACACHTCTGGGTCTGATCAAAAACCGAAGAAATACGTGCTTGGAGACACTCAT		730
Qy	589	TRPASNPRHEALASPRHEGLINTHRANLEUGLYILEARGVALASPRGLYASNLYS-LY		606
Db	731	TGGAAATTTGGCGATTTCATGACTGAACAGTCCACGACGAGAGTCTGGGGGAATGAAA		790
Qy	608	SGIYVALRPHETNARGASPRATGLYPRGLYALALALAHISERTLEUARGVALAAGTR		622
Db	791	GGGGAATCTTCACTCGGCACAGACCAACAAAAAGTCAGCGTTCTTTTGCAGAGAGATA		850
Qy	628	PTNRSERILEASPRYS 633		
Db	851	CTGGAGATTTTGCAA 866		

BM557676	1055 bp	mRNA	linear	EST 20-FEB-2002
LOCUS				
DEFINITION	AGENCOURT_6586423 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5477657			
ACCESSION	BM557676			
VERSION	5', mRNA sequence.			
KEYWORDS	BM557676.1 GI:18799844			
SOURCE	EST.			
ORGANISM	Homo sapiens (human)			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	1 (bases 1 to 1055)			
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgabbs-remail.nih.gov			
	Tissue Procurement: ATCC			
	cDNA Library Preparation: Rubin Laboratory			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)			
	DNA Sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov			
	Plate: LCM1996 row: a column: 18			
	High quality sequence stop: 668.			
FEATURES				
source				
	1..1055			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:5477657"			
	/tissue_type="astrocytoma grade IV, cell line"			
	/lab_host="DH10B (phage-resistant)"			
	/clone_id="NIH_MGC_98"			
	/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."			
ORIGIN				
Alignment Scores:				
Pred. No.:	4,34e-66	Length:	1055	
Percent Similarity:	689.00	Matches:	158	
Best Local Similarity:	53.08%	Conservative:	49	
Query Match:	40.51%	Mismatches:	114	
	20.54%	Indels:	70	
	4	Gaps:	11	
US-10-757-093-4 (1-634) x BM557676 (1-1055)				
QY	257	11eaapgiuaepgiyalaliallevalaialaalsasergiyalaglnglythrvalthrile	276	
	:::	:::	:::	
Db	3	TTGGATGCAGAAAACAAGTCGTGGCGGAATGGAACTGGACCCAGCCCACTTAAGTG	62	
QY	277	ProSeRValVLeuTfPInPro-----GlyAlaAlaTfLeuTfIn	291	
Db	63	CCAGGTGCACCCCTCTGTGGCCCGTACCTGATGCAGAACCCCTGCTACTATCTGATTC	122	
QY	292	LeuGInVAlaAnIleValGly-----SerSeRgiYaepValVAlaApThTfYasneu	309	
	:::	:::	:::	
Db	123	TTGGAGGGTGCAGCTGACTGCACAGACGTCACCTGGGGCCCTGTGTCTGACTTCTTCACTTC	182	
QY	310	AlaThrGlyValaArgThrValVlySeValaIaGlySerGInPheLeuIleAsnGlyVlySePro	329	
	:::	:::	:::	
Db	183	CCTGTGGGGATCCGCACTGTGGCTGTCAACCAAGCCAGCTTCTTATCAATGGGAACTT	242	
QY	330	PheTfYrPheThrGlyPheGlyVlyeHfIseGluApThrAlaValaIaRgGlyVlyeGlyVHfIaAp	349	

Db	243	TTCTATTTCACGGGTGTCACCAACGATGAGGATCGGACATCCGAGGAAAGGCGTTCCGAC	302
Oy	350	ProIaIyrmecValiHiIaBpPheGlnIleuMetIystrPileGlyAlaIaBserPhearg	369
Db	303	TGGCGCGCTGGCGTGAAGACTTCAACCTGCTGGCTGGCTGGCCAAAGCTTTCCGT	362
Oy	370	ThisSerhiStryProTyraIaGluGluValMetAserPheAlaAserPArgAserGlyIleVal	389
Db	363	ACGAGCCACTACCCCTATGACAGAGAGATGTGCAGATGTGTGACCGCTATGGAGATTGTG	422
Oy	390	ValIleAserGlnThrProAlaValAlGlyLeuAsnIleAlaIleuMetGlyValserIuSer	409
Db	423	GTCATTCAGATAGTGTCCCGCGCTGGCGCTGGCGCGCTG-----	458
Oy	410	GlyAlaProGlnThrPheThrProAserAlaIleAserAplystrhGlnGluAlaHiIeIyS	429
Db	459	-----CCGACGAGTCTTC-----AACAACGTTTCTCTGCATACCAACACATG	497
Oy	430	GlnAlaIleAserGluIleAlaAserPheValAserPheAlaAserValMetTrpSer	449
Db	498	CAGGTGTGTAAGAAAGTGTGCTGTAGGACAAAGAACCCCGCGGTGTATGTGGTCT	557
Oy	450	IleAlaAserGluProAlaAserHiIeGluAserGlyAlaArgGlyIyPheGluProleuthr	469
Db	558	GTGGCCAAACGAGCTGGCTGCCCACTAGAACTGTGGCTACTT-GAA-----	607
Oy	470	AenIeuthrArgIleuAserProThrArgProIleThrPheAlaAenValGlyThrAla	489
Db	608	-----GCT	610
Oy	490	ThrTyrlleuAserAArgIleAserAserPheAserValserCyBileAserAArgTyrrhe	509
Db	611	CCGAT-----GTGATGTGATCTGTGTGAACAGCTACTAC	646
Oy	510	GlyTrpTyrrserGlnThrGlyAserPheGluGluIleAglAlaIleAglIleGluIleu	529
Db	647	TCTTGATATCACGACTACGAGGACCTGTGATGATACGCTGCAGCTGCCACCCAGCTTT	706
Oy	530	HiIeGlyTrpGlnGlyIyPheHiIeAserProIleValMetThrGlyIyGlyAlaAserThr	549
Db	707	GAGAACTGGTATGAAGATACAGAAAGCCATTATTACAGAGGATATGAGAGCAAAAG	766
Oy	550	LeuAlaGlyLeuHiIeSerIleLeuGlyIleuProTrpSerGluGluPheGlnValGlnMet	569
Db	767	ATTCCAGAGGTTTCCACGAGATCCACTGTATGTTCACTGAAGAGTACCAANAGCTCG	826
Oy	570	LeuAserPheTyrr-----HiIeArgValPheAserArgIleGluSer	582
Db	827	CTAAGACAGTACCCATCTGGCTGTGATCAAAAACGCAAAATACTGGGTGTGAGAGC	886
Oy	583	MetAlaGlyGluHiIeValTrpAserPheAlaAserPheGlnThrAenLeuGlyIleIleArg	602
Db	887	TCTTTTGA-----AATTTTGCCCGATTTCAT---GACTTGACAGTACCCGA	931
Oy	603	ValI-----AserGlyAserIleLeuGlyValPhe---ThrArgAserArgIyAserPro	617
Db	932	ACAAAGTGCCTTGGAATTAATAAAAGGGGGAATTCCTCCGGCAGAGACAAACCA	991
Oy	618	LyAlaIaAlaIaHiIeSerLeuAArgAlaArg	627
Db	992	AAAAGTGCACGCCGTTTCTTTTTCGAAA	1021
RESULT 15			
CP406520			
LOCUS	CP406520	877 bp	mRNA linear EST 02-SEP-2003
DEFINITION	CH3#042_G0317 Canine heart normalized cDNA library in plnuescrpt		
ACCESSION	Canis familiaris cDNA clone CH3#042_G03 5', mRNA sequence.		
VERSION	CP406520.1 GI:34406534		
KEYWORDS	EST.		
SOURCE	Canis familiaris (dog)		
ORGANISM	Canis familiaris		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

